

Opinion piece



Cite this article: Maney DL. 2016 Perils and pitfalls of reporting sex differences. *Phil. Trans. R. Soc. B* **371**: 20150119. <http://dx.doi.org/10.1098/rstb.2015.0119>

Accepted: 15 November 2015

One contribution of 16 to a theme issue 'Multifaceted origins of sex differences in the brain'.

Subject Areas:

behaviour, neuroscience

Keywords:

gender differences, male, female,
sex differences in the brain,
US National Institutes of Health policy

Author for correspondence:

Donna L. Maney
e-mail: dmaney@emory.edu

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rstb.2015.0119> or via <http://rstb.royalsocietypublishing.org>.

Perils and pitfalls of reporting
sex differences

Donna L. Maney

Department of Psychology, Emory University, Atlanta, GA 30322, USA

The idea of sex differences in the brain both fascinates and inflames the public. As a result, the communication and public discussion of new findings is particularly vulnerable to logical leaps and pseudoscience. A new US National Institutes of Health policy to consider both sexes in almost all preclinical research will increase the number of reported sex differences and thus the risk that research in this important area will be misinterpreted and misrepresented. In this article, I consider ways in which we might reduce that risk, for example, by (i) employing statistical tests that reveal the *extent to which* sex explains variation, rather than *whether or not* the sexes 'differ', (ii) properly characterizing the frequency distributions of scores or dependent measures, which nearly always overlap, and (iii) avoiding speculative functional or evolutionary explanations for sex-based variation, which usually invoke logical fallacies and perpetuate sex stereotypes. Ultimately, the factor of sex should be viewed as an imperfect, temporary proxy for yet-unknown factors, such as hormones or sex-linked genes, that explain variation better than sex. As scientists, we should be interested in discovering and understanding the true sources of variation, which will be more informative in the development of clinical treatments.

1. Introduction

Sex differences in the brain have made headlines for more than a century. In 1912, James Crichton-Browne, a prominent neuropsychologist and collaborator of Darwin, explained in a *New York Times* article why 'women think quickly' and 'men are originators':

In woman, Sir James said, the posterior region of the brain receives a richer flow of arterial blood, in men the anterior region. The work of the two regions of the brain is different. The posterior region is mainly sensory and concerned with seeing and hearing. The anterior region includes the speech centre, the higher inhibitory centres, which are concerned with will, and the association centres, concerned with appetites and desires based upon internal sensations.

There is, Sir James thinks, a correspondence between the richer blood supply of the posterior region of the brain in women and their delicate powers of sensuous perception, rapidity of thought and emotional sensibility, and between the richer blood supply of the anterior region in men and their greater originality on higher levels of intellectual work, their calmer judgment and their stronger will [1, p. 4].

Although we may find such revelations archaic and even a bit offensive, the same type of thinking remains prevalent today. News reports and information-based websites such as Wikipedia, WebMD and HowStuffWorks.com contain an alarming amount of pseudoscience. It is commonly asserted, for example, that women listen with both sides of the brain, whereas men use only the left side [2,3] and that women use white matter to think, whereas men use grey [4,5]. Women allegedly have 10 times as much white matter as do men, whereas men have 6.5 times as much grey matter as do women ([6,7], reviewed in [8]). Whereas women navigate using cerebral cortex, men use 'an entirely different area' that is 'not activated in women's brains' [6]. Such assertions, although inaccurate, are easy to find on the Internet and in the popular press.

The misrepresentation of sex differences is likely to become even more commonplace. Partly because of increasing availability of imaging technologies, the percentage of journal articles that refer to sex differences and the brain has more than doubled in the past two decades (figure 1*a*). Over the same period, media

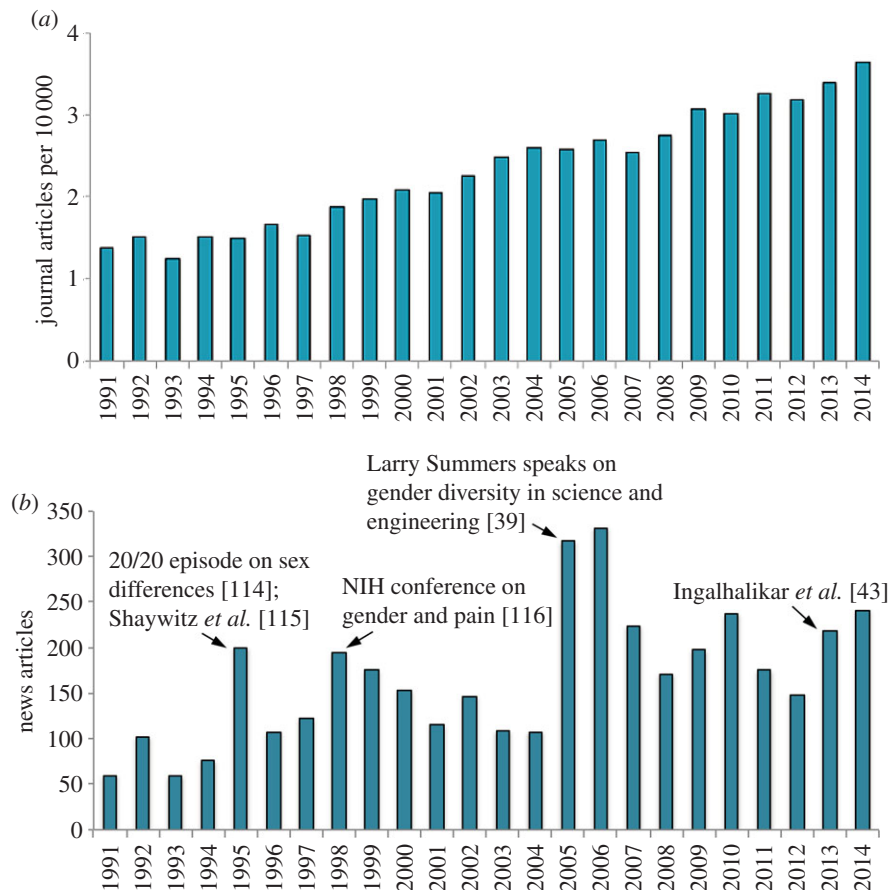


Figure 1. Interest in sex differences in the brain has increased over the past 25 years. (a) The percentage of journal articles that mention sex or gender differences and brain (Web of Science) has approximately doubled. (b) The number of news articles about sex differences in the brain (Proquest) has increased approximately fivefold. Arrows indicate years during which a particular study or event contributed substantially to an increase in news stories. For the methods and raw data, see the electronic supplementary material, table S1. (Online version in colour.)

reporting on the topic has risen by about fivefold (figure 1b). These increases are already impressive, but the amount of research on sex differences is about to increase even further, far beyond what figure 1 could foretell. This year, the US National Institutes of Health (NIH) mandated the inclusion of both sexes in most research with animals, tissues or cells [9]. The new policy requires NIH-funded researchers to disaggregate data by sex and, when possible, compare the sexes. The goal is to ‘transform how science is done’ [10]. Research on sex differences is thus set to expand from a small percentage of studies to nearly *all* studies funded by NIH. If this goal is achieved, the side effects will almost certainly include a fresh onslaught of questionable interpretations and claims.

In this article, I outline some traps that researchers face as they test for and report sex differences. These pitfalls are largely related to the interpretation of statistical tests, choice of wording and the use of inference. I suggest below a number of strategies that may help researchers avoid common problems and therefore minimize misinterpretation and misrepresentation of their work.

2. Three fallacies of sex differences

Miscommunication of the nature and meaning of sex differences can be traced to many causes [11–14]. Here, I outline three problematic ways of thinking, or fallacies, that have impeded the communication of findings. First, we usually present our conclusions about sex differences as the answer to a yes-or-no question when the real answer lies somewhere in

between ‘yes’ and ‘no’. Second, we often attempt to infer the behavioural or evolutionary function of a sex difference in the brain without sufficient evidence to do so. Third, we tend to assume that sex differences are caused by genetic or hormonal influences rather than by experience. At the root of all three of these points is a fourth issue, which perhaps could be regarded as a fourth fallacy: the notion that sex acts as an independent variable that measurably affects other variables. Sex is merely a label; defining it in biological terms has proven tricky [15]. Sex is at best a proxy for the more important and interesting factors that covary with sex [16].

(a) Fallacy 1: with respect to any trait, the sexes are either fundamentally different or they are the same

Sex has been called a basic biological variable that splits the population into two halves [17]; the categories of male and female are regarded as rigidly discrete [18,19], forming ‘taxa’ [14]. When these two populations are compared, however, measures of most traits overlap extensively [14,19–21]. The conceptualization and communication of that overlap are impeded by our natural urge to dichotomize [22] and by language choices that emphasize difference [12,23]. For example, if a statistical test returns a low p -value, we are likely to make statements such as, ‘females outperform males on the memory task’ or ‘women are more susceptible than men to the side effects’. Taken literally, these statements imply that with respect to the trait measured, males and females constitute distinct groups [24]. In nearly all cases, however, that interpretation is wildly incorrect. Conversely, when the p -value is not low enough to reject the null

hypothesis of sameness, we often conclude that the sexes are the same even when sex could explain some important variation [25–28]. The problem here is that we are asking a yes-or-no question when both ‘yes’ and ‘no’ are the wrong answer. To truly understand the nature of most sex differences, which arguably are not actual ‘differences’, we need to ask *how much* the sexes differ, not *whether or not* they do [25,26].

(b) Fallacy 2: the cause of a sex difference in behaviour or ability can be inferred from functional neuroanatomy

It is a longstanding tradition to invoke sex differences in neuroanatomy to explain alleged sex differences in behaviour, intellect or other traits [29]. Just as the New York Times printed in the early twentieth century that male-like patterns of blood flow allow more original thinking [1], in *this* century more white matter in women is said to confer greater language skills [30] and ability to multitask [7]. The larger hippocampus of women is said to support better memory [31], language skills [32], learning skills [33] and processing of emotive information [34]. The inferior parietal lobule, larger on the left in men and on the right in women [35], apparently underlies differences in math ability and sensitivity to crying babies [36]. Testosterone-induced lateralization of brain function is claimed to increase men’s interest in machines [37], while oestradiol increases women’s attention to emotions and communication [38]. Each of these anatomical or hormonal differences has been invoked to explain why men and women tend to enter different types of professions [39–41].

The above assertions are based on the following logic: (i) a structure (or hormone) we’ll call ‘X’ differs between men and women; (ii) X is related to a behaviour we’ll call ‘Y’; (iii) men and women differ in Y; therefore, the sex difference in X causes the sex difference in Y. This argument is invalid because it invokes the false cause fallacy—a sex difference in Y cannot be deduced to depend on X. In addition to being invalid, the argument is also often unsound in that rarely are all three premises supported. Evidence that structure X plays a role in behaviour Y, for example, is usually scarce. Even in animal models, in which lesions and other manipulations can be performed, the behavioural functions of sexually differentiated brain regions are, for the most part, unclear [42]. Evidence of a sex difference in behaviour Y is also sometimes lacking, and instead a stereotype is offered. Consider the following popular inference: (i) the hemispheres of the brain are more heavily interconnected in women than in men [43]; (ii) greater hemispheric interconnectedness allows better multitasking; (iii) women are better multitaskers than men, therefore the anatomical difference explains the difference in ability [7,44]. First, the evidence that variation in interhemispheric connections actually contributes to variation in human abilities is practically non-existent [45]. Second, studies of multitasking have shown no female advantage [46,47]. The argument pervades popular culture nonetheless, probably because it appears to confirm stereotypes [8,11,48].

(c) Fallacy 3: sex differences in the brain must be preprogrammed and fixed

A third type of fallacious thinking, which pervades news stories and scholarly articles alike, is signalled by words

such as ‘hardwired’, ‘natural’ and ‘genetic’. These terms are nearly always used to argue that sex stereotypes are rooted in biology. They make sex differences sound predetermined and inevitable, untouched by experience or culture [8,11,23,49,50]. Readers are easily convinced, particularly when the explanations appear to support their own biases [8,48]. Such arguments, however, ignore the exquisite plasticity of the brain; the effects of sex-linked genes and sex hormones on neuroanatomy are irreducibly entangled with the effects of sex-specific experience [23]. Suggesting otherwise leads to logical leaps known as the *appeal to nature* and *deterministic fallacies*—that sex-typical behaviour is natural, predetermined and out of our control. The cost of these fallacies is high: readers exposed to such arguments are more likely to endorse stereotypes and engage in stereotype-consistent behaviour (reviewed in [12,51,52]), and may feel powerless to change their own trajectories [48,50].

In §§3 and 4, I will focus primarily on Fallacy 1 and how to avoid it. The remaining two fallacies are important in the context of communicating findings in research papers as well as press releases and are revisited in §5.

3. Pink hippocampus, blue hippocampus? Most are purple

Becker *et al.* [53] defined a sex difference as a dimorphism, in other words a trait that ‘occurs in two forms, one form typical of males and the other typical of females.’ (p. 1651). The trouble with such definitions is that, except for sex chromosomes, gonads and external genitalia [54], the two sexes rarely take two distinct forms. The vast majority of sex differences in neuroanatomy and physiology are characterized by overlapping distributions. Variation that is attributable to sex can often require large sample sizes to detect. It is almost never the case that the sexes can be distinguished by a single structure in the brain that is said to ‘differ’. In other words, we typically cannot identify the sex of an individual by measuring any one thing in the brain; the majority of values fall into a grey area [14,19–21]. Conversely, values cannot be predicted accurately from the sex of an individual [21,54]. The overlap between the sexes is usually not represented clearly by scientists and conflicts with public perception of sex differences in the brain [8,11,49].

Figure 2 illustrates often-cited sex differences in humans. The graphs, which were made using the reported means and standard deviations (table S2), show the frequency distributions, or the number of individuals of each sex with any given measure or score. Note that Reis & Carothers [14] have shown that for many sex differences, actual data do not cluster according to sex as they do in figure 2, but rather fall onto a single continuum for both sexes.

Because most readers are familiar with the sex difference in human height, I have depicted that first [55], for comparison (figure 2*a*). This sex difference is relatively large, as is the difference in total brain volume (figure 2*b*) [56]. When brain volume is controlled, sex differences in individual brain regions are smaller or disappear completely [66]. One of the most often-cited sex differences in the brain is that of the hippocampus, which some authors have reported is larger in women than men [21]. A recent meta-analysis showed no sex difference in this structure [67], and even when such a difference has been detected [21,57] there is a good deal of overlap. In the dataset depicted in figure

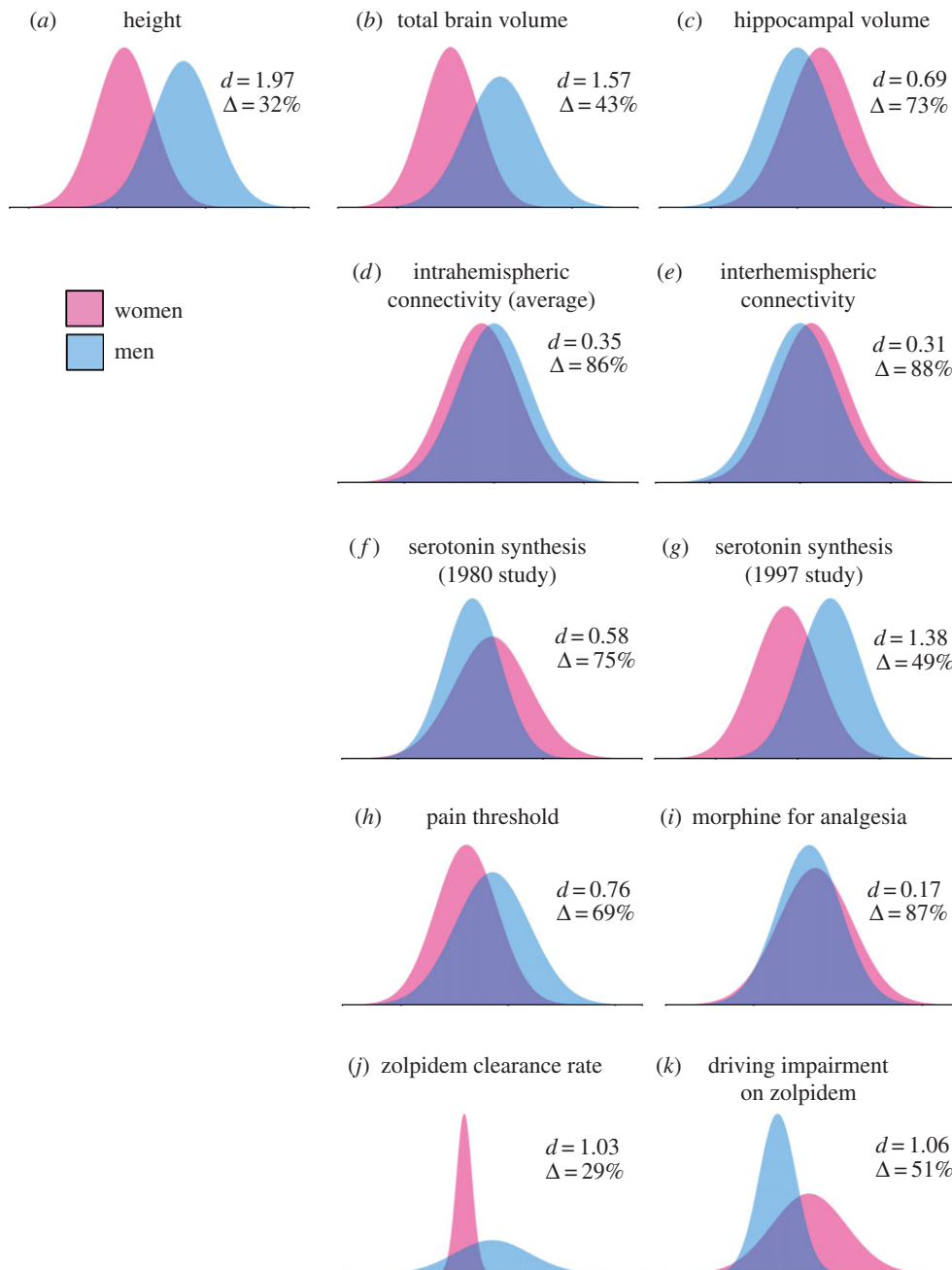


Figure 2. Some of the most-often cited sex differences in humans are characterized by extensive overlap (Δ). In each panel, d represents Cohen's d and Δ represents the percentage overlap. The graphs show the frequency distributions, or the number of individuals of each sex (y -axis) with any given measure or score (x -axis). (a) Distributions for human height [55] are shown for comparison. The effect size is large, but men and women overlap in height by 32%. (b) Total brain volume is larger in men than in women [56]. (c) The volume of the hippocampus, corrected for total brain volume, has been reported to be larger in women than men [21,57]. (d) Intraheispheric and (e) interheispheric connectivity, measured via diffusion tensor imaging, varies slightly according to sex [43]. The effect size plotted in (d) represents the average of 18 comparisons for which the values were higher in men. (f) Serotonin synthesis has been reported to be higher in women [58] and, (g) in a later study, higher in men [59]. (h) Pain thresholds are generally reported to be higher in men (data shown from [60]; see [61] for review). (i) The dosage of morphine required for analgesia may vary according to sex but the degree of overlap is high (data shown from [62]; see [63] for review). (j) The drug zolpidem, a popular sleep aid, is cleared by women more slowly than by men [64]. (k) The morning after taking zolpidem, women are more impaired than men during a driving task [65]. For the values used to make the plots, see the electronic supplementary material, table S2. Distributions were assumed normal in each case. All graphs were made using the interactive tool at www.sexdifference.org. Readers are encouraged to use this tool to assess overlap for sex differences that they find in the literature or in their own research.

2c, hippocampal size is more typical of the opposite sex, i.e. it is larger than average in men or smaller in women, in about a third of the population. Thus, despite a statistically significant sex difference ($p < 0.0001$) we cannot say that the hippocampus occurs in two forms.

The claim of two forms is commonly made, even when the extent of overlap is quite large. In the paper on interheispheric connectivity in humans cited in §2, Ingallhalikar *et al.* [43]

concluded that there are 'fundamental differences' between male and female brains. The authors did not report the degree of overlap, but the reported T statistics and degrees of freedom indicate that the sexes overlapped by nearly 90% (figure 2d,e; see also [68]). Nonetheless, the study was hailed by both scientists and the media as strong evidence that male and female brains take two distinct forms [13,69]. News stories announced that 'men's brains go back to front, women's go side

to side' [70] and that the structural differences are 'so profound that men and women might almost be separate species' [71].

Sex differences are sometimes interpreted as evidence that, despite overlap, an entire sex is somehow deficient. For example, early findings of a lower rate of serotonin synthesis in men (e.g. [58]) (figure 2f) have been used in the popular press to argue that a serotonin deficit makes men impulsive and 'stupid' [72]. This sort of interpretation of a sex difference, in other words that one sex exhibits a deficit, can lead to singling out of one sex for interventions. The alleged serotonin deficit in men, for example, has been argued by educators to warrant specialized educational strategies for boys [33]. In 1997, a different study [59] suggested that serotonin synthesis may actually be higher in men (figure 2g). This finding eventually led to a reversal in the popular press such that women became the abnormal sex. In a book on women's mental health, Dr David Edelberg [73, p. 14] wrote, 'When doctors discovered the relationship between low serotonin and emotional disorders, they started comparing the serotonin levels between sexes and found that women simply were not making enough'. As more clinically relevant biomarkers are discovered to vary according to sex, presenting and emphasizing overlap between the sexes should help prevent the impression that an entire sex is atypical.

The main goal of the new NIH policy [9] is to balance our approach to understanding diseases and clinical conditions, the aetiology of which may vary according to sex [10]. An often-highlighted example of such a condition is pain. Of the dozens of reported sex differences in pain, two with among the largest sample sizes are shown in figure 2. Figure 2*h* shows lower pain thresholds for women than men—which is typical not only for pressure [60], but also thermal, electrical and other types of pain [61]. Sex differences in the response to analgesia, on the other hand (figure 2*i*), are characterized by greater overlap [62] and less agreement about which sex exhibits the greater response. Some authors have reported that opioid analgesics are more effective in women than men, some the other way around (reviewed in [63]). Although animal research has suggested that mechanisms of morphine analgesia differs between males and females [74], sex differences in morphine efficacy have been difficult to detect in humans—they may be masked by differences in side effects or pain thresholds [62,63,74].

Perhaps the most popular example of a drug with differential effects in men and women is zolpidem, the sleep aid in Ambien. Even after controlling for body mass, the clearance rate of this drug is lower in women than in men [64] (figure 2*j*), which has clear consequences. The morning after taking zolpidem, for example, women deviated more from a straight line while driving—in other words, they were more impaired (figure 2*k*) [65]. The U.S. Food and Drug Administration (FDA) recently issued new guidelines reducing the dosage for women [75], which was hailed by advocates of personalized medicine as a huge step in the right direction. Cahill [76] pointed out that 'millions of women had been overdosing on Ambien'. That is most certainly the case. Note, however, that a sizeable proportion of the men fall into the female range for clearance rate (figure 2*k*). If most of the women taking Ambien were overdosing, then about a third of the men were doing the same. In their 2013 announcement [75], the FDA actually recommended lower doses for both sexes. The guidelines stated, 'These lower doses of zolpidem will be effective in most women *and many men*' (p. 3; italics added). Despite the overlap acknowledged by the FDA, the change in

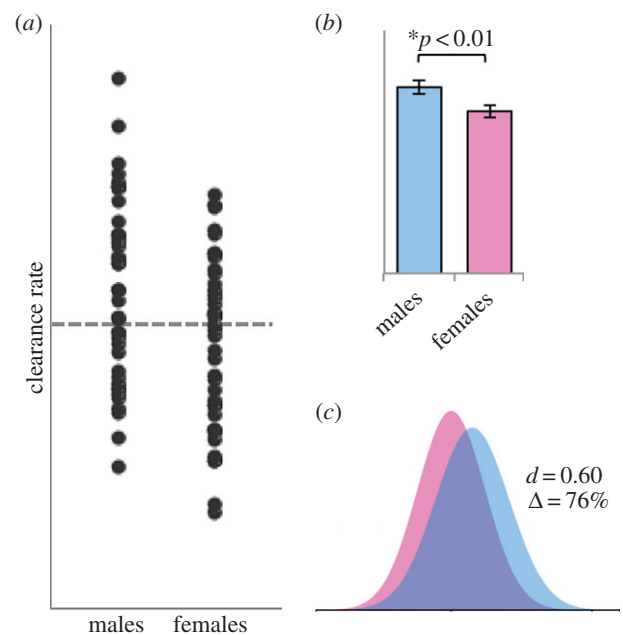


Figure 3. Analysis of a hypothetical dataset for 'Dimorphinil', a fictional drug. Normally distributed data were artificially generated for 40 'individuals' of each sex. (a) The values both above and below the population mean (grey dotted line) consist of 18 members of one sex and 22 of the other. In other words, the numbers of above- and below-average individuals are about the same for each sex. (b) A Student's *t*-test shows a significant sex difference ($p < 0.01$). (c) The effect is medium-sized ($d = 0.60$) but there is a high degree of overlap ($\Delta = 76\%$). For the methods and raw data, see the electronic supplementary material, table S3.

guidelines for zolpidem remains by far the most-cited example of the need for sex-specific medicine.

A statistically significant sex difference does not necessarily indicate a meaningful separation between the sexes. Figure 3 shows hypothetical data for a fictional drug I will call 'Dimorphinil'. In this fictive sample of 40 men and 40 women, the sex difference in clearance rate is both significant ($p < 0.01$) and of medium size ($d = 0.60$). The effect size exceeds that for some real drugs, for example, cyclosporine and nifedipine, for which clearance rate differs significantly between the sexes [77,78]. The low *p*-value and medium effect size for Dimorphinil suggest non-trivial, clinically relevant sex difference [79]. Yet, the percentage of males and females above and below the overall mean is about the same—22 of the males and 18 of the females cleared Dimorphinil faster than the overall average and 18 of the males and 22 of the females cleared it more slowly. If we were to recommend different dosages of Dimorphinil for men and women, only a handful of patients would benefit—those in the tails of the distributions, which consist mostly of one sex. The majority of patients, for whom sex does not reliably predict clearance rate, could be harmed by too much or too little drug. Sex-based dosages, even in this case, may be preferable to a one-size-fits-all approach, if only because benefiting a small minority of patients is nonetheless a benefit. It is critical to remember, however, that when significant sex differences exist, they usually indicate that sex explains some small portion of the variation, not that the sexes are 'different'. Treatments that are 'personalized' for each sex will benefit everyone only when the effect size is astronomical. Although some patients will certainly benefit from sex-specific medicine, it comes with a high price tag: it over-emphasizes

difference and strengthens false notions that men and women fall into dichotomous categories of patients.

Overlap between the sexes indicates that other factors, in addition to sex, contribute to variation in a trait. Because sex is itself not a mechanism [16] and cannot be absolutely defined in biological terms [15], it is at best a proxy for these other variables. Many of them are not yet known. Most sex differences, if they have not already, will eventually be completely explained by some other factor that covaries with sex. These explanatory factors will, I believe in every case, contribute more to our understanding of mechanism than does the label 'sex'. Take, for example, a study that showed a male advantage in multitasking [46]. The results conflicted with the popular notion that women are better multitaskers than men [7]. The more interesting result, however, was that the sex difference was completely explained by a much larger sex difference in video game experience. In other words, multitasking ability was actually predicted by video game experience, not sex. In this case, investigating the potentially explanatory correlates of sex was more informative and satisfying than simply reporting a sex difference. In the case of drug development, a better strategy than dividing patients by sex, or an important next step, would be to discover and study the covariates that explain sex differences. For example, let us imagine that unbeknownst to researchers, the sex difference in Dimorphinil clearance rate (figure 3) is explained by physical activity, which also depends on sex [80,81]. If we did not know about the effect of activity on Dimorphinil clearance, we might tailor dosage according to sex instead—female athletes and male couch potatoes would receive the wrong dose.

The list of known variables that can covary with sex is long [82], and includes some rather uninteresting, non-biological factors such as housing arrangements in animal facilities [83]. Perhaps for that reason, it is commonly argued that sex differences driven by obvious or uninteresting covariates, for example, body mass, are not 'true' sex differences [82]. If they are not caused by biological factors that covary with sex, what, then, are true sex differences? Those caused by sex hormones? Levels of sex steroids can overlap extensively, depending on the species and stage of development. Are true sex differences caused, then, by the sex-determining region of the Y chromosome? Some women have that gene [84]. As we peel away and discard each of the mechanisms that actually do explain sex differences, we are left only with the 'essence' of male and female; in other words, slippery concepts that gender scholars identify as the basis of 'essentialist' thinking [45,85]. Essentialism is not useful to us as neuroscientists [23,86]. In neuroscience, sex can be a *predictor* but never a *cause* [16]. Sex differences are all caused by knowable factors that covary with sex—those factors are not likely to be defined by sex. Our task is to discover and understand those factors, not simply to demonstrate that the sexes are different. The mechanisms underlying sex-based variation are incredibly complex, so for now, using sex as a proxy for the more interesting variables will have to suffice. Because sex is highly politicized and poorly communicated to the public, however, it is not a good stopping point.

4. Is there a difference? Both 'yes' and 'no' are wrong answers

When we compare the sexes, we divide our sample into two categories. Thus, the very nature of the scientific question

encourages us to think dichotomously. To make matters worse, the most commonly used statistical tests encourage dichotomous thinking about the results [25]. Our decision to declare the sexes 'different' or the 'same' is usually based on whether a p -value is above or below 0.05. But $p = 0.04$ and $p = 0.06$ represent essentially the same result and cannot lead logically to opposite conclusions. Further, no matter what our decision, we are almost certainly wrong. As noted above, finding a statistically significant difference does not mean that the sexes are substantively 'different'—the differences are almost always characterized by important overlap (figures 2 and 3). Even a strikingly low p -value may not indicate a meaningful difference; if the sample size is large enough, a statistically significant sex difference could be detected in any measure simply owing to noise [26,27].

Conversely, a p -value above the threshold for a significant difference (usually $p > 0.05$) does not indicate that the sexes are the same [25–27]. In other words, failure to reject the null does not give license to accept the null. In so doing, we would be accepting absence of evidence as evidence of absence. Cumming has called this logical error the 'fallacy of the slippery slope of non-significance' [25]. The error is particularly common in psychology and neuroscience, fields that rely heavily on null hypothesis significance testing [27]. Hoekstra *et al.* [28] found that in the field of psychology, 60% of authors concluded 'no difference' when $p > 0.05$. If one's goal is to confirm that sex does not matter for the measurement at hand, the t -test would seem a rather poor choice to test for sameness.

Rather than asking a yes-or-no question about whether the sexes differ, it is more informative to quantify the extent to which, or *how much*, sex contributes to variation [25]. The p -value obtained from null hypothesis significance tests, e.g. t -tests, does not answer this question. Lower p -values do not indicate larger effects. Measures of effect size, such as Cohen's d , are more useful (figure 2). When d is less than about 0.5, regardless of statistical significance, sex is unlikely to explain important variation and the finding should probably not be emphasized without good reason [79]. Even if no significant difference is detected, reporting the effect size is helpful to determine next steps [23]. Estimates of confidence intervals [25,27] or Bayesian approaches [87] represent other alternatives to null hypothesis significance testing. Overlap or similarity between groups can be estimated [88,89]. The companion web page of this article, www.sexdifference.org, is an online tool that calculates effect size and percentage overlap from user-entered descriptive statistics. It can be used to visualize distributions of the user's own data or, as was done in figure 2, those of published sex differences.

Whether we are calculating p -values or effect sizes, to detect a sex difference we must compare the sexes directly. Disaggregation of data by sex, now mandated by NIH, does not involve an actual comparison. We can simply test for an effect of treatment in each sex independently. When the p -value is below alpha for one sex but not the other, we typically claim a 'sex-specific effect'. Such conclusions are problematic for many reasons. As critics of the NIH policy have pointed out [87,90], dividing a sample into subgroups lessens power and therefore the ability to detect effects. In a famous illustration of this phenomenon, Sleight [91] recounted the analysis of data from the International Study of Infarct Survival, which showed a clear benefit of daily aspirin. When the population was divided by astrological

sign, resulting in 12 separate subgroups, the beneficial effect of aspirin was lost in the Libras and Geminis. Testing a hypothesis within each sex incurs a similar risk that an effect will be detected in one sex but not the other, when in fact both sexes are responding.

The problem with conducting independent tests in males and females goes beyond the loss of statistical power. Such a design does not actually allow us to test whether the effect of treatment depends on sex. To answer that question, we must test for *interactions* between sex and treatment. Nonetheless, the practice of testing two groups independently is quite common in neuroscience. In an analysis of articles in top-ranking journals such as *Nature Neuroscience* and *Neuron*, Nieuwenhuis *et al.* [92] found that authors tested for an interaction in only half of the cases in which two experimental effects were compared. In the rest, the authors tested for effects in the two groups independently. When $p < 0.05$ for one group and $p > 0.05$ for the other, they concluded that the effect of treatment depended on the group. This conclusion is simply another case of the slippery slope of non-significance [25]. Upon failing to reject the null for one group, the authors accepted the null—and worse, contrasted the result with that of the other group. But when the p -values of two tests differ, the outcomes themselves cannot be said to differ [27,92,93].

Testing whether experimental effects differ between two groups requires a test for interactions, for example, factorial analysis of variance (ANOVA). An advantage of ANOVA is that the main hypothesis can be tested using the entire sample of males and females together, thus avoiding the Libra/Gemini problem described above [91]. Although power to detect an interaction is notoriously low in ANOVA [94], a low-powered test is perhaps acceptable in the context of developing sex-specific medical treatments because we are interested in detecting only robust, clinically relevant interactions. Note, however, that if we fail to detect an interaction, particularly with a low-powered test, we cannot say that the sexes respond in the same way to treatment—in so doing we once again slide down the slippery slope [25].

5. Communicating sex differences

What is the best way to plot a sex difference? The most valuable representations of our data will accurately depict overlap. Plotting the distributions (figure 2; see www.sexdifference.org) allows the reader to assess the effect size. Other options to depict overlap include graphing confidence intervals using error bars or ‘cat’s eye pictures’ [95]. Alternatively, plotting individual data points, for example, in dot plots (figure 3), allows the reader to see exactly the extent of overlap as well as the percentages of each sex in the tails of the distributions. Many readers, particularly those outside the field, may look only at the figures; thus depicting overlap graphically likely pays off even more than adjustments to language in the paper.

Once the graphs are made and the paper accepted, some of the most important work is yet to be done. Publication of newsworthy results is generally accompanied by a press release, in most cases written by staff in the public relations department of the home institution. This press release, much more than the paper itself, sets the tone of media coverage and dictates the information contained therein. Even journalists who specialize in science writing may skip the journal article and read only the press release [96]. Although scientists may blame the media

for misrepresenting and sensationalizing their findings [97–100], press releases often contain the same oversimplification, omission of information and misinterpretations as do news stories [13,51,52,97,101]. The process of packaging our findings into sound bites naturally leads us into the three traps outlined above in §2. Difference is more popular than sameness [11,102], which may compel us to downplay overlap and announce striking sex differences—thus invoking Fallacy 1.

In order to make newly discovered sex differences more meaningful to the public, researchers are prone to speculate about the functions of those differences—thus invoking Fallacy 2. In an analysis of a highly covered study on hemispheric connectivity (mentioned above; figure 2*d,e* [43]), O’Connor & Joffe [13] found that dubious claims in the news stories were actually present in the press release. Some of the claims could be traced to the journal article itself. Although the study contained no behavioural data, the authors listed a number of sex differences in domains such as memory, social cognition and sensorimotor skills that their data might explain. The press release highlighted these domains and introduced new ones, such as multitasking. All of these were picked up by the popular media, which added even more alleged sex differences supposedly explained by the findings. The small sex effects on connectivity detected in the study (figure 2*d,e*) were said to explain sex differences in emotional intelligence, intuition, athleticism, hunting, cleaning the house and endless other so-called gendered behaviours [13,71]. A typical headline stated, ‘The hardwired difference between male and female brains could explain why men are ‘better at map reading’ [103].

By using the term ‘hardwired’, that headline also invoked Fallacy 3: the idea that sex differences are predetermined and immutable. The issue of predeterminism is particularly relevant to results from animal models, which are often extrapolated to humans more hastily than is warranted. We are told both by media and by researchers that because a sex difference appears in a non-human animal, it must be genetic, shaped by evolution and free of sociocultural influence. Take, for example, a recent paper by Farmer *et al.* [104]. The authors showed that in mice, experiencing pain caused females to spend less time with males. By contrast, males in pain continued to pursue females. In their paper, the authors argued that the findings ‘suggest that the well-known context sensitivity of the *human* female libido can be explained by evolutionary rather than sociocultural factors.’ (p. 5747; italics added). The press release [105], which led with ‘Not tonight honey...’ proved irresistible; the study received worldwide coverage. The fact that it was conducted in mice, not humans, was sometimes lost, however. Headlines announced, ‘Women’s low sex drive due to biological reaction to pain’ [106] and ‘It’s science: why your headache excuse is actually legit but his isn’t’ [107]. The coverage of this study shows clearly that animal studies are not immune to widespread media attention or potential over-interpretation.

Although the authors of a study do not typically write the press release, they are often quoted and may even be given opportunities to edit. Thoughtful attention, particularly to how the work will be interpreted by the public, is important at this stage. The press release should be regarded as a ‘point of no return’; once unleashed, misinformation evolves on its own and can be difficult, if not impossible, to rein in [13,98]. Rather than offering questionable functional or evolutionary explanations for the results, it may be more effective to point out what the research does *not* indicate. In a recent press release from Northwestern University [108,109], the senior author

emphasized that research on sex differences in the brain ‘is not about things such as who is better at reading a map or why more men than women choose to enter certain professions’. Rather, the author emphasized that it is ‘about making biology and medicine relevant to everyone, to both men and women’. The ensuing media coverage was both widespread and higher in quality than the usual. Clearly, close collaboration with the PR department can pay off. Authors may even want to take the lead in communicating findings to the public; more and more scientists use social media and blogs for this purpose [96,100]. Ultimately, although we all want to share our findings widely and ponder what they mean, the most effective communications—those that enhance public understanding—will stick to the facts and avoid speculations about evolutionary function or hardwiring. The topic of sex differences is too volatile and easily sensationalized to risk doing otherwise.

6. Future steps and an alternative ending

The inclusion of both sexes in biomedical research is a necessary and important step forward. Comparing the sexes, and their responses to potential treatments, will inform not only the development of inclusive medicine but also our understanding of mechanisms that covary with sex [110]. The new NIH policy [9] will advance these goals, but carries with it high risk of collateral damage. As more and more sex differences are discovered, the number of misinterpretations will also increase. It would be best to be prepared, ideally by providing training to researchers and journalists. The NIH Office of Research for Women’s Health already offers helpful online courses that cover the importance of studying both sexes, the biology of sexual differentiation and disorders that affect one sex disproportionately [82]. This training does not, however, cover how to recognize and avoid contributing to pseudoscience. In fact, the training materials currently state that ‘women have larger left cortical language receptors than men’ (course 2, lesson 4, p. 6) but the source cited [111] does not mention language. According to the same course, men mount a ‘fight or flight’ response when faced with a crisis whereas women ‘tend and

befriend’; this difference is said to be caused by a sex difference in oxytocin. One of the cited sources does not mention sex differences [112]; another states emphatically that the role of oxytocin in human behaviour is not well understood [113]. Certainly, if these sorts of misrepresentations can creep into NIH training materials, we can expect them to pop up practically anywhere—including in our own work [13,52]. Training in the interpretation and communication of sex differences should be a priority.

In this article, I have painted a rather grim picture of fallacious headlines appearing in the news every day. Whether the new NIH guidelines actually increase the number of such headlines depends, of course, not only on the manner in which results are communicated, but also on the extent to which the guidelines are followed. Depending on how stringently they are enforced, there could be a very different but equally disappointing outcome. Research is expensive, and not all researchers are interested in sex differences. When asked to increase sample sizes and perform additional analyses, a majority of researchers may be motivated to rule out sex differences as quickly as possible. In such cases, demonstrating that the sexes are the same becomes highly incentivized. After a cursory *t*-test showing ‘no difference’, researchers may feel free to go back to business as usual. Thus, whereas dubious interpretations of positive findings threaten public scientific literacy, false negatives may turn out to be much more detrimental to the mission of the NIH. For every sex difference that makes headlines, numerous others may go undiscovered as they slip down the slippery slope of non-significance and out of sight.

Data accessibility. The datasets supporting the figures have been uploaded as the electronic supplementary material.

Competing interests. I have no competing interests.

Funding. My effort devoted to writing this article was supported by Emory University.

Acknowledgements. I am grateful to Evan Goode, who designed and implemented the interactive tool at sexdifference.org. I also thank Chris Goode, Kim Wallen and Catherine Woolley for helpful comments on the manuscript.

References

1. Anon. 1912 Sex differences in brain: Crichton-Browne tells why women think quickly, men are originators. *New York Times* 10 May 1912, p. 4.
2. Anon. 2013 Five amazing things women’s bodies can do. *Cosmopolitan*. See <http://www.cosmopolitan.com.au/health-lifestyle/feel-good-advice/2013/10/five-amaze-things-womens-bodies-can-do> (accessed 18 September 2015).
3. Anon. 2015 Sex differences in humans. Wikipedia. See https://en.wikipedia.org/w/index.php?title=Sex_differences_in_humans&oldid=677415914 (accessed 22 August 2015).
4. Carey B. 2005 Men and women really do think differently. *Live Science*. See <http://www.livescience.com/3808-men-women-differently.html> (accessed 7 August 2015).
5. Edmonds M. 2008 Do men and women have different brains? How Stuff Works. See <http://science.howstuffworks.com/life/inside-the-mind/human-brain/men-women-different-brains1.htm> (accessed 6 August 2015).
6. 2015 How male and female brains differ. WebMD. See <http://www.webmd.com/balance/features/how-male-female-brains-differ> (accessed 8 August 2015).
7. CBS News 2010 Size matters: how male, female brains compare. See <http://www.cbsnews.com/news/size-matters-how-male-female-brains-compare> (accessed 18 June 2013).
8. Maney DL. 2014 Just like a circus: the public consumption of sex differences. *Curr. Topics Behav. Neurosci.* **19**, 279–296. (doi:10.1007/7854_2014_339)
9. 2015 *Consideration of sex as a biological variable in NIH-funded research*. National Institutes of Health Notice Number: NOT-OD-15–102, issued 6-9-2015.
10. Clayton JA, Collins FS. 2014 NIH to balance sex in cell and animal studies. *Nature* **509**, 282–283. (doi:10.1038/509282a)
11. Eliot L. 2011 The trouble with sex differences. *Neuron* **72**, 895–898. (doi:10.1016/j.neuron.2011.12.001)
12. Joel D, Kaiser A, Richardson SS, Ritz SA, Roy D, Subramaniam B. 2015 A discussion on experiments and experimentation: NIH to balance sex in cell and animal studies. *Catalyst: Feminism, Theory, Technoscience*, 1.1.
13. O’Connor C, Joffe H. 2014 Gender on the brain: a case study of science communication in the new media environment. *PLoS ONE* **9**, e110830. (doi:10.1371/journal.pone.0110830)
14. Reis HT, Carothers BJ. 2014 Black and white or shades of gray: are gender differences categorical or dimensional? *Curr. Dir. Psychol. Sci.* **23**, 19–26. (doi:10.1177/0963721413504105)
15. Ainsworth C. 2015 Sex redefined. *Nature* **518**, 288–291. (doi:10.1038/518288a)
16. Springer KW, Stellman JM, Jordan-Young RM. 2012 Beyond a catalogue of differences: a theoretical

- frame and good practice guidelines for researching sex/gender in human health. *Soc. Sci. Med.* **74**, 1817–1824. (doi:10.1016/j.socscimed.2011.05.033)
17. Miller VM. 2014 Why are sex and gender important to basic physiology and translational and individualized medicine? *Am. J. Physiol. Heart Circ. Physiol.* **306**, H781–H788. (doi:10.1152/ajpheart.00994.2013)
 18. Haslam N, Rothschild L. 2000 Essentialist beliefs about social categories. *Brit. J. Soc. Psych.* **39**, 113–127. (doi:10.1348/014466600164363)
 19. Carothers BJ, Reis HT. 2013 Men and women are from Earth: examining the latent structure of gender. *J. Pers. Soc. Psychol.* **104**, 385–407. (doi:10.1037/a0030437)
 20. Hyde JS. 2014 Gender similarities and differences. *Annu. Rev. Psychol.* **65**, 373–398. (doi:10.1146/annurev-psych-010213-115057)
 21. Joel D. *et al.* 2015 Sex beyond the genitalia: the human brain mosaic. *Proc. Natl Acad. Sci. USA* **112**, 15 468–15 473. (doi:10.1073/pnas.1509654112)
 22. Dawkins R. 2011 The tyranny of the discontinuous mind. *The New Statesman*. See <http://www.newstatesman.com/blogs/the-staggers/2011/12/issue-essay-line-dawkins> (accessed 29 September 2015).
 23. Rippon G, Jordan-Young R, Kaiser A, Fine C. 2014 Recommendations for sex/gender neuroimaging research: key principles and implications for research design, analysis, and interpretation. *Front. Hum. Neurosci.* **8**, 650. (doi:10.3389/fnhum.2014.00650)
 24. Cimpian A, Brandon AC, Gelman SA. 2010 Generic statements require little evidence for acceptance but have powerful implications. *Cog. Sci.* **34**, 1452–1482. (doi:10.1111/j.1551-6709.2010.01126.x)
 25. Cumming G. 2012 *Understanding the new statistics: effect sizes, confidence intervals, and meta-analyses*. New York, NY: Routledge.
 26. Kline RB. 2004 *Beyond significance testing: reforming data analysis methods in behavioral research*. Washington, DC: APA Books.
 27. Fine C, Fidler F. 2015 Sex and power: why sex/gender neuroscience should motivate statistical reform. In *Handbook of neuroethics*, pp. 1447–1462. The Netherlands: Springer.
 28. Hoekstra R, Finch S, Kiers H, Johnson A. 2006 Probability as certainty: dichotomous thinking and the misuse of *p*-values. *Psychol. Bull. Rev.* **13**, 1033–1037. (doi:10.3758/BF03213921)
 29. Shields S. 1975 Functionalism, Darwinism, and the psychology of women. *Amer. Psychol.* **30**, 739–754. (doi:10.1037/h0076948)
 30. Wolpert L. 2014 *Why can't a woman be more like a man? The evolution of sex and gender*. New York, NY: Skyhorse.
 31. LaMotte S. 2015 Men's memories worse than women's, especially with age. CNN. See <http://www.cnn.com/2015/03/17/health/male-memory/> (accessed 12 August 2015).
 32. Gross G. 2014 How boys and girls learn differently. *Huffington Post*. See http://www.huffingtonpost.com/dr-gail-gross/how-boys-and-girls-learn-differently_b_5339567.html (accessed 5 November 2015).
 33. Gurian M, Stevens K. 2004 With boys and girls in mind. *Educ. Leadership* **61**, 21–27.
 34. Jantz GL. 2014 Brain differences between genders. *Psychology Today*. See <https://www.psychologytoday.com/blog/hope-relationships/201402/brain-differences-between-genders> (accessed 28 October 2015).
 35. Frederikse ME, Lu A, Aylward E, Barta P, Pearson G. 1999 Sex differences in the inferior parietal lobule. *Cereb. Cortex* **9**, 896–901. (doi:10.1093/cercor/9.8.896)
 36. Hensley A. 2009 10 big differences between men's and women's brains. Masters of Health Care. See <http://www.mastersofhealthcare.com/blog/2009/10-big-differences-between-mens-and-womens-brains/> (accessed 29 September 2015).
 37. Baron-Cohen S. 2003 *The essential difference: the truth about the male and female brain*. New York, NY: Basic Books.
 38. Brizendine L. 2007 *The female brain*. New York, NY: Broadway Books.
 39. Summers L. 2005 Faculty diversity: research agenda. Paper presented at the Science and Engineering Workforce Project meeting 24 January 2005, National Bureau of Economic Research, Cambridge, MA.
 40. Lukas C. 2008 Studying women and science: why women's lower rate of participation in science, technology, engineering, and mathematics courses isn't a problem for the government to solve. Independent Women's Forum, Position Paper #8.
 41. Sommers CH. 2008 *Why can't a woman be more like a man?* American Enterprise Institute. See <https://www.aei.org/publication/why-cant-a-woman-be-more-like-a-man-3/> (accessed 29 September 2015).
 42. de Vries GJ, Södersten P. 2009 Sex differences in the brain: the relation between structure and function. *Horm. Behav.* **55**, 589–596. (doi:10.1016/j.yhbeh.2009.03.012)
 43. Ingallhalikar M *et al.* 2014 Sex differences in the structural connectome of the human brain. *Proc. Natl Acad. Sci. USA* **111**, 823–828. (doi:10.1073/pnas.1316909110)
 44. 2013 Brain connectivity study reveals striking differences between men and women. Penn Medicine. See http://www.uphs.upenn.edu/news/news_releases/2013/12/verma/ (accessed 10 September 2015).
 45. Fine C. 2010 *Delusions of gender*. New York, NY: Norton.
 46. Hambrick DZ, Oswald FL, Darowski ES, Rench TA, Brou R. 2010 Predictors of multitasking performance in a synthetic work paradigm. *Cogn. Psychol.* **24**, 1149–1167. (doi:10.1002/acp.1624)
 47. Mäntylä T. 2013 Gender differences in multitasking reflect spatial ability. *Psychol. Sci.* **24**, 514–520. (doi:10.1177/0956797612459660)
 48. Morton TA, Postmes T, Haslam SA, Hornsey MJ. 2009 Theorizing gender in the face of social change: is there anything essential about essentialism? *J. Pers. Soc. Psychol.* **96**, 653–664. (doi:10.1037/a0012966)
 49. Fine C. 2014 His brain, her brain? *Science* **346**, 915–916. (doi:10.1126/science.1262061)
 50. Brechman JM, Lee CJ, Cappella JN. 2011 Distorting genetic research about cancer: from bench science to press release to published news. *J. Commun.* **61**, 496–513. (doi:10.1111/j.1460-2466.2011.01550.x)
 51. Fine C. 2012 Explaining, or sustaining, the status quo? The potentially self-fulfilling effects of 'hardwired' accounts of sex differences. *Neuroethics* **5**, 285–294. (doi:10.1007/s12152-011-9118-4)
 52. Fine C. 2013 Neurosexism in functional neuroimaging: from scanner to pseudo-science to psyche. In *The Sage handbook of gender and psychology* (eds M Ryan, N Branscombe), pp. 45–61. London, UK: Sage.
 53. Becker JB *et al.* 2005 Strategies and methods for research on sex differences in brain and behavior. *Endocrinology* **146**, 1650–1673. (doi:10.1210/en.2004-1142)
 54. Joel D. 2011 Male or female? Brains are intersex. *Front. Integr. Neurosci.* **5**, 57. (doi:10.3389/fnint.2011.00057)
 55. Sperrin M, Marshall AD, Higgins V, Renehan AG, Buchan IE. In press. Body mass index relates weight to height differently in women and older adults: serial cross-sectional surveys in England (1992–2011). *J. Pub. Health*. (doi:10.1093/pubmed/fdv067)
 56. Li W, Tol MJ, Li M, Miao W, Jiao Y, Heinze HJ, Bogerts B, He H, Walter M. 2014 Regional specificity of sex effects on subcortical volumes across the lifespan in healthy aging. *Hum. Brain Map.* **35**, 238–247. (doi:10.1002/hbm.22168)
 57. Biswal BB *et al.* 2010 Toward discovery science of human brain function. *Proc. Natl Acad. Sci. USA* **107**, 4734–4739. (doi:10.1073/pnas.0911855107)
 58. Young SN, Gauthier S, Anderson GM, Purdy WC. 1980 Tryptophan, 5-hydroxyindoleacetic acid and indoleacetic acid in human cerebrospinal fluid: interrelationships and the influence of age, sex, epilepsy and anticonvulsant drugs. *J. Neurol. Neurosurg. Psych.* **43**, 438–445. (doi:10.1136/jnnp.43.5.438)
 59. Nishizawa S, Benkelfat C, Young SN, Leyton M, Mzengeza SD, De Montigny C, Blier P, Diksic M. 1997 Differences between males and females in rates of serotonin synthesis in human brain. *Proc. Natl Acad. Sci. USA* **94**, 5308–5313. (doi:10.1073/pnas.94.10.5308)
 60. Chesterton LS, Barlas P, Foster NE, Baxter GD, Wright CC. 2003 Gender differences in pressure pain threshold in healthy humans. *Pain* **101**, 259–266. (doi:10.1016/S0304-3959(02)00330-5)
 61. Fillingim RB, King CD, Ribeiro-Dasilva MC, Rahim-Williams B, Riley JL. 2009 Sex, gender, and pain: a review of recent clinical and experimental findings. *J. Pain* **10**, 447–485. (doi:10.1016/j.jpain.2008.12.001)
 62. Aubrun F, Salvi N, Coriat P, Riou B. 2005 Sex- and age-related differences in morphine requirements for postoperative pain relief. *Anesthesiology* **103**, 156–160. (doi:10.1097/0000542-200507000-00023)
 63. Niesters M, Dahan A, Kest B, Zacny J, Stijnen T, Aarts L, Sarton E. 2010 Do sex differences exist in opioid analgesia? A systematic review and meta-analysis of human experimental and clinical studies. *Pain* **151**, 61–68. (doi:10.1016/j.pain.2010.06.012)
 64. Greenblatt DJ, Harmatz JS, von Moltke LL, Wright CE, Duro ALB, Harrel-Joseph LM, Shader RI. 2000 Comparative kinetics and response to the benzodiazepine agonists triazolam and zolpidem:

- evaluation of sex-dependent differences. *J. Pharmacol. Exp. Ther.* **293**, 435–443.
65. Verster JC, Roth T. 2012 Gender differences in highway driving performance after administration of sleep medication: a review of the literature. *Traffic Injury Prev.* **13**, 286–292. (doi:10.1080/15389588.2011.652751)
 66. Jäncke L, Mérillat S, Liem F, Hänggi J. 2015 Brain size, sex, and the aging brain. *Hum. Brain Map.* **36**, 150–169. (doi:10.1002/hbm.22619)
 67. Tan A, Ma W, Vira A, Marwha D, Eliot L. 2015 The human hippocampus is not sexually dimorphic: meta-analysis of structural MRI volumes. *NeuroImage*. **124**, 350–366. (doi:10.1016/j.neuroimage.2015.08.050)
 68. Ridgeway G. 2013 Illustrative effect sizes for sex differences. Figshare. See http://figshare.com/articles/Illustrative_effect_sizes_for_sex_differences/866802 (accessed 21 December 2013).
 69. Cahill L. 2014 Fundamental sex difference in human brain architecture. *Proc. Natl Acad. Sci. USA* **111**, 577–578. (doi:10.1073/pnas.1320954111)
 70. Boesveld S. 2013 Men's brains go back to front, women's go side to side. *National Post* (Canada) 4 December 2013, p. 1.
 71. Naish J. 2013 Men's and women's brains: the truth! *Daily Mail*. See <http://www.dailymail.co.uk/femail/article-2518327/Mens-womens-brains-truth-As-research-proves-sexes-brains-ARE-wired-differently-womens-cleverer-ounce-ounce-men-read-female-feelings.html> (accessed 6 August 2015).
 72. Moir A, Moir B. 1999 *Why men don't iron*. New York, NY: Citadel Press.
 73. Edelberg D. 2006 *The triple whammy cure: the breakthrough women's health program for feeling good again in 3 weeks*. New York, NY: Free Press.
 74. Loyd DR, Murphy AZ. 2014 The neuroanatomy of sexual dimorphism in opioid analgesia. *Exp. Neurol.* **259**, 57–63. (doi:10.1016/j.expneurol.2014.04.004)
 75. U.S. Food and Drug Administration. 2013 Risk of next-morning impairment after use of insomnia drugs; FDA requires lower recommended doses for certain drugs containing zolpidem (Ambien, Ambien CR, Edluar, and Zolpimist). Drug Safety Communication issued 1 October 2013.
 76. Cahill L. 2014 Equal \neq the same: sex differences in the human brain. *Cerebrum*. See http://dana.org/Cerebrum/2014/Equal_The_Same_Sex_Differences_in_the_Human_Brain/ (accessed 21 December 2015).
 77. Kahan BD, Kramer WG, Wideman C, Flückner SM, Lorber MI, van Buren CT. 1986 Demographic factors affecting the pharmacokinetics of cyclosporine estimated by radioimmunoassay. *Transplantation* **41**, 459–464. (doi:10.1097/00007890-198604000-00009)
 78. Krecic-Shepard ME, Park K, Barnas C, Slimko J, Kerwin DR, Schwartz JB. 2000 Race and sex influence clearance of nifedipine: results of a population study. *Clin. Pharmacol. Ther.* **68**, 130–142. (doi:10.1067/mcp.2000.108678)
 79. Klein SL *et al.* 2015 Opinion: Sex inclusion in basic research drives discovery. *Proc. Natl Acad. Sci. USA* **112**, 5257–5258. (doi:10.1073/pnas.1502843112)
 80. Caspersen CJ, Pereira MA, Curran KM. 2000 Changes in physical activity patterns in the United States, by sex and cross-sectional age. *Med. Sci. Sports Exerc.* **32**, 1601–1609. (doi:10.1097/00005768-200009000-00013)
 81. Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M. 2008 Physical activity in the United States measured by accelerometer. *Med. Sci. Sports Exercise* **40**, 181. (doi:10.1249/mss.0b013e31815a51b3)
 82. The Science of Sex and Gender in Human Health: Online Course Site. National Institutes of Health, Office of Research on Women's Health. See Sexandgender.course.od.nih.gov (accessed 5 August 2015).
 83. Ritz SA *et al.* 2014 First steps for integrating sex and gender considerations into basic experimental biomedical research. *FASEB J.* **28**, 4–13. (doi:10.1096/fj.13-233395)
 84. Griffin JE, Wilson JD. 1980 The syndromes of androgen resistance. *New Engl. J. Med.* **302**, 198–209. (doi:10.1056/NEJM198001243020404)
 85. Eliot L. 2009 *Pink brain, blue brain*. Boston, MA: Houghton Mifflin Harcourt.
 86. Fine C, Duke R. 2015 Expanding the role of gender essentialism in the single-sex education debate: a commentary on Liben. *Sex Roles* **72**, 427–433. (doi:10.1007/s11199-015-0474-0)
 87. IOM (Institute of Medicine) 2012 *Sex-specific reporting of scientific research: a workshop summary*. Washington, DC: National Academies Press.
 88. Weitzman MS. 1970 Measures of overlap of income distributions of white and Negro families in the United States. Technical paper No. 22. Department of Commerce, Bureau of Census, Washington DC.
 89. Nelson JA. 2014 Are women really more risk-averse than men? A re-analysis of the literature using expanded methods. *J. Econ. Surv.* **29**, 566–585. (doi:10.1111/joes.12069)
 90. Fields RD. 2014 NIH policy: mandate goes too far. *Nature* **510**, 340. (doi:10.1038/510340a)
 91. Sleight P. 2000. Debate: subgroup analyses in clinical trials: fun to look at but don't believe them! *Curr. Control. Trials Cardiovasc. Med.* **1**, 25–27. (doi:10.1186/CVM-1-1-025)
 92. Nieuwenhuis S, Forstmann BU, Wagenmakers EJ. 2011 Erroneous analyses of interactions in neuroscience: a problem of significance. *Nat. Neurosci.* **14**, 1105–1107. (doi:10.1038/nn.2886)
 93. Gelman A, Stern H. 2006 The difference between 'significant' and 'not significant' is not itself statistically significant. *Amer. Stat.* **60**, 328–331. (doi:10.1198/000313006X152649)
 94. Wahlsten D. 1990 Insensitivity of the analysis of variance to heredity-environment interaction. *Behav. Brain Sci.* **13**, 109–120. (doi:10.1017/S0140525X00077797)
 95. Cumming G. 2014 The new statistics: why and how. *Psychol. Sci.* **25**, 7–29. (doi:10.1177/0956797613504966)
 96. Brumfiel G. 2009 Supplanting the old media? *Nature* **458**, 274–277. (doi:10.1038/458274a)
 97. Brown P. 2012 Nothing but the truth. Are the media as bad at communicating science as scientists fear? *EMBO Rep.* **13**, 964–967. (doi:10.1038/embor.2012.147)
 98. Mehr SA. 2015 Miscommunication of science: music cognition research in the popular press. *Front. Psychol.* **6**, 988. (doi:10.3389/fpsyg.2015.00988)
 99. Moore B, Singletary M. 1985 Scientific sources' perception of network news accuracy. *J. Quarterly* **62**, 816–823. (doi:10.1177/107769908506200415)
 100. Pew Research Center 2015 How scientists engage the public. See <http://www.pewinternet.org/2015/02/15/how-scientists-engage-public/> (accessed 6 August 2015).
 101. Eliot L. 2013 Sex-trapology in the latest brain science. *Huffington Post*. See http://www.huffingtonpost.com/lise-eliot/media-hype-and-the-scienc_b_4458458.html (accessed 27 Nov 2015).
 102. Richardson SS. 2014 Y all the hype? Slate. See http://www.slate.com/articles/double_x/doublex/2014/05/sex_difference_findings_in_human_genome_studies_be_very_skeptical.html (accessed 15 August 2015).
 103. Connor S. 2013 The hardwired difference between male and female brains could explain why men are 'better at map reading'. *The Independent*. See <http://www.independent.co.uk/life-style/the-hardwired-difference-between-male-and-female-brains-could-explain-why-men-are-better-at-map-reading-8978248.html> (accessed 12 August 2015).
 104. Farmer MA *et al.* 2014 Pain reduces sexual motivation in female but not male mice. *J. Neurosci.* **34**, 5747–5753. (doi:10.1523/JNEUROSCI.5337-13.2014)
 105. Chipello C. 2014 *Pain curbs sex drive in female mice, but not in males*. McGill University. See <https://www.mcgill.ca/channels/news/pain-curbs-sex-drive-female-mice-not-males-235376> (accessed 10 September 2015).
 106. Rose T. 2014 Women's low sex drive due to biological reaction to pain. *Guardian Liberty Voice*. See <http://guardianlv.com/2014/04/womens-low-sex-drive-due-to-biological-reaction-to-pain/> (accessed 10 September 2015).
 107. Chatel A. 2014 It's science: why your 'headache' excuse is actually legit. *Your Tango*. See <https://www.yourtango.com/2014214244/new-study-women-pain-tolerance-affects-sex-drive> (accessed 10 September 2015).
 108. Fellman M. 2015 Scientists uncover a difference between the sexes. Northwestern University. See <http://www.northwestern.edu/newscenter/stories/2015/08/scientists-uncover-a-difference-between-the-sexes.html> (accessed 5 September 2015).
 109. Tabatadze N, Huang G, May RM, Jain A, Woolley CS. 2015 Sex differences in molecular signaling at inhibitory synapses in the hippocampus. *J. Neurosci.* **35**, 11 252–11 265. (doi:10.1523/JNEUROSCI.1067-15.2015)
 110. McCarthy MM, Arnold AP, Ball GF, Blaustein JD, De Vries GJ. 2012 Sex differences in the brain: the not so inconvenient truth. *J. Neurosci.* **32**, 2241–2247. (doi:10.1523/JNEUROSCI.5372-11.2012)
 111. Gur RE, Kohler C, Turetsky BI, Siegel SJ, Kanes SJ, Bilker WB *et al.* 2004 A Sexually dimorphic ratio of orbitofrontal to amygdala volume is altered in schizophrenia. *Biol. Psychol.* **55**, 512–517. (doi:10.1016/j.biopsycho.2003.10.009)
 112. McCarthy MM, Altemus M. 1997 Central nervous system actions of oxytocin and

- modulation of behavior in humans. *Mol. Med. Today* **3**, 269–275. (doi:10.1016/S1357-4310(97)01058-7)
113. Taylor SE, Klein LC, Lewis BP, Gruenewald TL, Gurung RAR, Updegraff JA. 2000 Biobehavioral responses to stress in females: tend-and-befriend, not fight-or-flight. *Psychol. Rev.* **107**, 411–429. (doi:10.1037/0033-295X.107.3.411)
114. Stossel J. 1995 Boys and girls are different: men, women, and the sex difference. 20/20 (television programme), American Broadcasting Company.
115. Shaywitz BA *et al.* 1995 Sex differences in the functional organization of the brain for language. *Nature* **373**, 607–609. (doi:10.1038/373607a0)
116. NIH Pain Research Consortium 1998 Conference on gender and pain: a focus on how pain impacts women differently than men. April 1998.