Are Women Really More Risk-Averse than Men?

Julie A. Nelson
September 2012

Tufts University
Medford MA 02155, USA
http://ase.tufts.edu/gdae

View the complete list of working papers on our website:
http://www.ase.tufts.edu/gdae/publications/working_papers/index.html

Abstract:

While a substantial literature in economics and finance has concluded that women are more risk averse than men, this conclusion merits reconsideration. Drawing on literatures in statistics and cognitive science, this essay discusses the important difference between drawing conclusions based on statistical inference, which concerns aggregates such as mean scores, and generalization, which posits characteristics of individuals classified into kinds. To supplement findings of statistical significance, quantitative measures of substantive difference (*Cohen's d*) and overlap (the *Index of Similarity*) are computed from the data on men, women, and risk used in 28 published articles. The results are considerably more mixed and overlapping than might be expected. Paying attention to empirical evidence that challenges subjective cultural beliefs about sex and risk has implications for labor economics, finance, and the economics of climate change.

Are Women Really More Risk-Averse than Men?

Julie A. Nelson[1]

The Issue

"We find that women are indeed more risk averse than men" conclude economists Rachel Croson and Uri Gneezy (2009, 448) in their *Journal of Economic Literature* review article "Gender Differences in Preferences." They base their conclusion on a number of empirical studies that found statistically significant differences between men's and women's behavior, on average, in lottery experiments or investment strategies—and in fact very similar statements are endemic in the literature.[2] But consider these two statements:

A. *"In our sample, we found a statistically significant difference in mean risk aversion between men and women, with women on average being more risk averse."*

B. *"Women are more risk averse than men."*

While the two statements are often taken as meaning the same thing, there is in fact a wide gulf of meaning between statement A, which is a narrow statement that can be factually correct within the confines of a particular study, and statement B which is a broad statement that implies a stable characteristic of people according to their sex.

Distinguishing between the two statements is important, since perceptions of sex differences in risk-taking have become part of both public and academic discussions about financial market stability (e.g., Kristof 2009); labor market, business, and investment success (e.g., Eckel and Grossman 2008, 14; Booth and Nolen 2012, F56); and environmental policy (e.g., Kahan, Braman et al. 2007). Broad statements about sex differences are also commonly made regarding other behaviors, such as competitiveness or management style. On a methodological plane, understanding the distinction between Statement A and Statement B is, in fact, important for understanding what we can really know based on statistical inference, in any application.

The current essays explores this gulf primarily from a statistical point of view, but also touches on the linguistic, cognitive, epistemological, ideological, and policy-making causes and implications of sliding between two statements such as A and B. It is assumed

---

[2] Statements of the form "women are more risk averse than men" occur in, for example, Arano, Parker et al. (2010), Bernasek and Shwiff (2001), Booth and Nolen (2012), Borghans, Golsteyn et al. (2009)—and nearly every other article on risk in the reference list for this essay.

that one goal of an economic science that aspires to objectivity is to make sure that the inferences drawn from statistical studies reflect only the statistical results, and not the influence of subjective factors such as individually-held preconceptions or culture-specific beliefs. The essay shows how the accuracy and objectivity of economic research could be improved, both by expanding our toolbox of quantitative methods and by improving the precision with which we interpret and communicate the results of our econometric research projects.

What does "Women are more risk averse than men" mean?

What might the statement that "women are more risk averse than men" mean? While it generally appears in the literature as though it is summarizing empirical results, recent research in cognitive science indicates that the relationship between statistical inference and what is communicated by this phrase is not at all simple or direct.

The Statement Is Often Understood as Universal or Generic

Taken as a bald statement, the statement might be taken as indicating something that is *universally* true for every individual member of the classes "women" and "men." In this case, it would have to be true that every individual woman is more risk averse than every individual man. This exceedingly strong implication is not likely intended by those who write such statements, since just one example of a cautious man and a bold woman disproves it. Cognitive research, however (reviewed below), suggests that many end-users of research may understand the statement in this way.

A second possibility is that the statement is intended as what psychologists, linguists, and philosophers call a *generic noun phrase* (Gelman 2005, 3). A generic noun phrase is a generalization that "refers to a category rather than a set of individuals" and "express[es] essential qualities" implying that a category "is coherent and permits categorywide inferences" (Gelman 2005, 3). In a generic of the form "Fs are G," that is, "one is saying *of a kind of thing*, specified in the statement, that its members are, or are disposed to be G (or to [do] G) *by virtue of being of that kind*. The speaker conveys that being G is somehow rooted in what it is to be an F: G-ing is what Fs do (or are disposed to do) by virtue of being F" (Haslanger 2011, 13). Examples include "tigers have stripes." In this case, a few counterexamples—e.g., albino tigers—do not nullify the statement, since having stripes is considered to be part of the intrinsic, essential nature of tigerhood, even if stripes are not manifested in a particular case (Leslie forthcoming). "We essentialize a kind if we form the (tacit) belief that there is some hidden, non-obvious, and persistent property or underlying nature shared by members of that kind, which causally grounds their common properties and dispositions," writes Leslie (forthcoming). Generics seem to "articulate core conceptual beliefs" (Khemlani, Leslie et al. 2012, 1).

In the current example, the statement would imply that greater risk-aversion is an essential characteristic of womanliness—or, by parallel reasoning, that greater risk-

seeking is an essential characteristic of manliness.[3] Such an essentializing view is evident in the risk and gender literature in article titles such as "Will Women Be Women?" (Beckmann and Menkhoff 2008) and "Girls will be Girls" (Lindquist and Säve-Söderbergh 2011). The apparent presumption in such titles is that were a group of women or girls found to *not* be relatively more risk averse, they would somehow be abnormal relative to their own female natures. Many articles treat risk-aversion as a sex-linked "trait" (e.g., Powell and Ansic 1997; Borghans, Golsteyn et al. 2009)—presumably stable across time and cultural contexts. In addition, many studies hypothesize evolutionary explanations for female risk aversion or male risk-seeking (e.g., Olsen and Cox 2001; Hartog, Ferrer-i-Carbonell et al. 2002; Cross, Copping et al. 2011), or hypothesize links to sex-related hormones or other genetic factors commonly thought to define an essence of femaleness or maleness (see examples cited in Meier-Pesti and Penz 2008; Croson and Gneezy 2009). The belief that there is an "innate, genetic, or biological basis" for the statement is also a characteristic of many generics (Gelman 2005, 1).

Generics Convey More (or Less) Than Statistical Prevalence

Rather than being intended as universal or generic, a third possibility is that the general and qualitative statement "women are more risk averse than men" is simply meant as a convenient shorthand for the more nuanced and accurate statement that "women exhibit, on average, more risk-averse behavior than men in this particular study." One might suppose that the generic statement "Fs are G" might be functionally equivalent to quantitative statements such as "Fs are *more* G, *on average*," "*most* Fs are G," or "*a majority of* Fs are G." This turns out not to be the case, however, since the relationship between statistical prevalence and core beliefs expressed as generics is far from simple or direct.

The more nuanced quantitative statement (e.g. "Fs are *more* G, *on average*") has been called an "*aggregate*-type proposition" in the statistical literature (Bakan 1955; Bakan 1966, 433). Unlike a generic, it implicitly acknowledges that not all women or all men are the same. Those who understand statistics will grant (at least upon reflection) that this means that the distributions of a characteristic among women and among men could overlap. The statement of an aggregate does not (logically) imply any story about causations or essences. An aggregate sort of statement can be empirically supported, within the context of a particular study. It expresses particular statistical results, not broad core beliefs.[4]

The complicated relation between generic statements and quantitative statements has been examined in the psychological and philosophical literatures. While some generic statements seem to be accepted as true based on statistical prevalence (e.g., "cars have

---

[3] The philosophical and linguistic literatures actually discuss various kinds of generics. For example, the generic statement "Mosquitos carry the West Nile virus" is generally considered true even though only a tiny minority of mosquitos carry the virus, due to the characteristic of carrying a disease being striking and dangerous (Leslie 2008). The essentialist version of a generic, however, is the one most relevant for the case examined here.

[4] In the articles reviewed below, some examples of such more nuanced (and accurate) presentations of results can be found, even though generic statements predominate.

radios" discussed in Khemlani, Leslie et al. (2012)), and some research has suggested Bayesian models for the formation of generics (Khemlani, Leslie et al. 2012, 10), statistical prevalence cannot, in fact, explain many cases. Prevalence in a majority of a kind seems to be neither necessary nor sufficient for a generic to be considered true. For example, the generic statement "Ducks lay eggs" is generally accepted as true. In fact, only a minority of ducks (i.e., those that are female, mature, and non-sterile) lay eggs. We seem to reason that ducks are birds, and birds, as a "kind" or category, have the "characteristic" of reproducing by way of eggs (Khemlani, Leslie et al. 2012). On the other hand, "Canadians are right-handed" is rejected, even though a majority of Canadians are right-handed (Khemlani, Leslie et al. 2012).

Rather than theorizing that generics are based in statistical regularities, then, a more promising line of research investigates the role of inductive reasoning and generalization in providing efficient (if far from fail-proof) modes of mental processing, as part of the process of human cognitive development (Gelman 2005; Leslie 2008). Empirical studies suggest that from a very early age, children create simple mental categories and classifications according to presumed essences (Leslie 2008). This propensity is so strong and basic that even among adults a statement phrased as a generic and accepted as true predisposes people to believe that *individual members* of a class will have the stated property (Khemlani, Leslie et al. 2009, 447)—that is, to interpret generic statements in ways that essentialize or even universalize the association. For example, given the statement "Quacky is a duck," people tend to agree with the statement "Quacky lays eggs" (Khemlani, Leslie et al. 2012, 10).

One can also note that—tempting as it may be to think so—the truth of generics cannot be proven through statistical analysis. That is, evidence of an association between F (a kind) and G (a characteristic) in a statistical study does not provide an adequate basis for asserting that G is of the essence of F-ness. "Essences" and "kinds" are simply not the sorts of things that can be empirically observed in the outside world. Rather, the tendency to class things in terms of kinds seem to be part of the structure of our *inside* worlds— that is, of evolved, developmental human cognition. Correlations between biological genetic, hormonal, or evolutionary phenomena and differences in behavior by sex, for example, can be *suggestive of* "essential" differences, but are not conclusive evidence of essential differences. "Essential differences" may, for example, be only one among a number of possible explanations: Perhaps a third, confounding variable underlies the observed relationship between F and G. In regard to perceived differences by sex, for example, pressure to conform to prescribed social roles or relative positions in hierarchies of power have frequently been suggested as confounding variables (Acker 1990; Kimmel 2000; Barnett and Rivers 2004; Hyde 2005; Eliot 2009; Fine 2010).

Another possibility is that further evidence may reveal that the behavior is not as closely tied to "kinds" as was earlier believed: Perhaps a broader sample of F will reveal evidence of not-G (e.g., the "black swan" observed after inductive reasoning led to the belief that swans are white, as discussed in Taleb (2010)). In the present context, for example, cross-cultural examination of sex differences in behavior may call beliefs about essences into question (Gneezy, Leonard et al. 2009; Blau, Ferber et al. 2010, 357; Else-

Quest, Hyde et al. 2010; Henrich, Heine et al. 2010). The fact that one may find explanations from biological or evolutionary "essences" plausible or even compelling does not mean that these explanations are true: Cognitive scientists use the term "confirmation bias" to refer to the human tendency to more readily absorb information that conforms to one's pre-existing beliefs.

It has been suggested (Bakan 1966) that perhaps some confusion among researchers between inductive reasoning and Fisherian statistical inference may be behind considerable misinterpretation of statistical results. To reason inductively means to go from specific observations to hypothesizing general—i.e., universal or generic—propositions that invite conclusions about individuals or kinds. Fisherian inference, on the other hand, means going from sample results concerning a aggregate, such as a sample difference in means, to inferences about the corresponding population aggregate. Fisherian significance testing about a difference in means, for example, only (at best) justifies the inference that a difference in means in a sample corresponds to a *difference in means* in a population. It does not justify generalizing an (sample) aggregate statement to a generic or universal statement.

Such issues have also been pointed out in the broader economics literature. Deaton's (Deaton 2009, 27-30) discussion of randomized control experiments in development economics makes a similar point: When all we get from a statistical study (no matter how well-designed) is information about differences in means, this does not justify inferences about other aspects of the distributions; does not supply the causal stories relating the variables of interest; and does not let us make predictions about individual events. Only when we *read into* the statistical results the existence of a generic relationship—e.g., "dams harm development" or "women are more risk averse"—are we tempted to, for example, make predictions about the next woman or the next dam.

When Is a Difference of Substantive Importance?

The implicit message behind a communicated finding of "difference"—especially if it is marked as the primary finding of a published article—is that the difference found is of a such a large substantive size that the result is worth reporting. As has been discussed to some extent in the economics literature, "substantive significance" and "statistical significance" are two different things (Ziliak and McCloskey 2004; Miller and Rodgers 2008). As is well known, differences can be statistically significant (i.e., generalizable to a larger population, under classical hypothesis testing) even if of small absolute size, especially in large samples. In the gender-and-risk literature, as in other literatures, however, judgments of "significant difference" are generally based on statistical significance alone. Discussions of the absolute size of the difference, much less its possible implications for society or policy, are rare.[5] This is in notable contrast to the psychological literature on behavior, where the expression of findings in terms of substantive difference is widespread, standardized, considered best practice (Wilkinson and Task Force on Statistical Inference 1999)—and where the implications for the study

---

[5] In the literature reviewed here, Eckel and Grossmann (2008, 15) and Dhomen, Falk, et al (2011, 530, 540) are notable for providing any extended discussion of substantive economic significance.

of sex differences has been the topic of intense professional discussion (Eagly 1995; Hyde and Plant 1995; Archer 1996; Martell, Lane et al. 1996, and other articles in the March 1995 and February 1996 issues of the American Psychologist.).

The default assumption that the difference must have been worth reporting, along with the abovementioned human tendency to essentialize or universalize based on generic statements, can be expected to contribute to the formation of (perhaps unfounded) beliefs that differences are substantively important, categorical, and possibly even non-overlapping. Anecdotal evidence from this author's undergraduate classes suggests that statements such as "women are more risk averse than men" tend to often be understood as sharply disjunctive, along the lines of "women are risk averse and men are not." This seems to be especially true the closer the "essential" characteristic is associated with a biological factor: Students in these classes have been (at first response) unanimous and adamant in believing that the men and women have distributions of "the number of children given birth to" that are strictly non-overlapping. Presumably, if the above-mentioned cognitive studies are correct, this is because the cognitively-available generic statement "Women bear children" tends to be more readily cognitively accessed than the empirical observation that some women, like men, never bear a child. Some of this tendency to neglect overlap can also be discerned in the risk literature. For example, as will be discussed below, two articles looking at financial risk concluded that females as a group have such a different attitude towards risk from men as a group, that they should always be matched with female financial advisors. The tendency to attribute dramatically distinct "essences" to men and women—so disjunctive as to make the two sexes as different as two different species—has been nicknamed the "Mars-versus-Venus" phenomenon, after the best-selling popular book *Men are from Mars, Women are from Venus* (Gray 1993). Examples of such assertions in the popular literature are now legion (see examples listed in Fine 2010).

Substantive size, however, is eminently measurable and can be made a direct object of discussion for statistical researchers, and potentially (though perhaps with more effort) for the larger public.

"Difference" and the Culture of Publication

A further complication in interpreting this literature comes from the first phrase in the more precise formulation, "*In our sample*, we found a statistically significant difference in mean risk aversion between men and women…"
Hypotheses about sex and risk aversion have clearly now been tested many times in the published literature—and, importantly, also in research findings that have remained unpublished.

This raises the issue of what is known in meta-analytical statistics as the "file drawer effect." When a particular hypothesis is tested many times, statistically significant results will sometimes appear by chance even if a true difference is nonexistent or very weak (e.g. in the pure classical hypothesis testing case, in 10 studies out of 100 when $\alpha=.10$). It is also quite clear that a statistically significant result is considered more

publishable than the finding of a lack of statistical significance. This biases the literature as a whole towards the reporting of difference, and against the reporting of similarity (as briefly mentioned in Croson and Gneezy 2009, 468). While the present study did not attempt to dig into file drawers, the analysis below suggests some publication bias towards the reporting of statistically significant differences only, even *within* the published works reviewed.

Caution is Warranted

In summary, the idea that a generic statement can serve as an innocuous shorthand for the accurate presentation of a study's results must be questioned, in the light of findings from cognitive science. Humans have a strong disposition to interpret generics as universals. Whether or not an economist who states "women are more risk averse than men" believes that there is some "essence" being communicated, the bald statement does much more than simply save the space that would have been used by inserting the qualifier, "on average": It tends to create an expectation among readers that that an individual woman can be predicted to be relatively risk-averse, and an individual man can be predicted to be relatively risk-loving. Other terms for this sort of false generalization are, of course, "stereotyping" and "prejudice."

Given the human tendency to infer essences or universals from generic statements, a responsible researcher who wants to accurately present the results of his or her study, then, would be wise to confine himself or herself to precisely stated aggregate propositions, and discuss substantive as well as statistical significance. At a larger disciplinary level, these tendencies also suggest that reviewers and editors as well as authors should not overlook findings of similarity, if the creation of biases and distortions in disseminated knowledge is to be avoided.

To be clear, the current essay is not arguing that "there is no difference between men and women," nor—by any means—does it take any stand concerning the existence or non-existence of metaphysical "essences." The concern here is with making the communication of empirical results conform to what can legitimately be inferred from the data. For accomplishing this, some expansion of economists' usual statistical toolbox may be helpful.

Statistical Tools for Investigating Sameness and Difference

While no set of summary statistics can give a complete picture of the details of similarities and differences between two distributions, the statistics leading to the qualitative judgment of "statistically significant or not" could, for a start, be supplemented by quantitative measures of substantive significance. This section describes two such statistics, which are then, in the next section, used to give an enriched perspective on 28 published articles that address issues of sex and risk.

The first measure is *Cohen's d*, a measure of difference in terms of "effect size" that is already in very wide use in the psychology, education, and neuropsychology literatures. The second is what will be called the *Index of Similarity*. This measure is newly proposed in the current essay, but is derived from measures already in use in labor and housing research.

Cohen's d

Cohen's d is one measure of "effect size," expressing the magnitude of a difference between means (e.g., Byrnes, Miller et al. 1999; Wilkinson and Task Force on Statistical Inference 1999; Hyde 2005; Cross, Copping et al. 2011). A difference may most directly, of course, be expressed in the same units as the underlying variable—as a number of dollars, bets, or units on a Likert scale (when such scales are treated as cardinal), etc. But such an expression of difference has two drawbacks. First, as has been stressed in much of the psychology literature, it cannot be easily compared across studies since it is not in standardized units. Secondly, it gives little insight into the substantive significance of the difference. Without knowing how much variation there is *within* groups, there is no way of knowing whether a between-group difference in means expressed in natural units implies a trivial or a huge divergence in behavior *between* the groups.

Cohen's d goes some way towards relieving these problems by expressing the difference between means in standard deviation units. For the case of a male versus female comparison, it is conventionally calculated as

$$d = \frac{\overline{X}_m - \overline{X}_f}{s_p}$$

where $\overline{X}_m$ is the male mean, $\overline{X}_f$ is the female mean, and s_p is the pooled standard deviation, a measure of the average within-group variation.[6] As conventionally set up in the psychological literature on gender differences, a positive value for *d* represents a case where the male score exceeds the female score. The difference is now expressed in standardized (standard deviation) units, and the measure quite sensibly gives a reduced measure of "difference" as the within-group variability (reflected in a rising s_p) increases.

---

[6] This is most often estimated as:

$$s_p = \sqrt{\frac{(n_m - 1)s_m^2 + (n_f - 1)s_f^2}{n_m + n_f}}$$

where s_m, s_f, n_m and n_f are the standard deviations and sample sizes for the male and female samples. While this seems to be the most common formula used in the psychology and education literatures, slightly different alternative formulations have also been proposed (e.g., Zakzanis 2001). Econometricians may find an opportunity to make contributions in this area, since some of the existing discussions seem to be weak on statistical theory—for example, Durlak (2009, 923) suggests guidleines that misinterpret the meaning of confidence intervals.

Cohen's d can, in theory, take on values from −∞ to +∞, with the extremes occurring the case where *all* women but *no* men (or vice versa) share exactly the same particular characteristic (so that the numerator has a nonzero value and there is simultaneously no within-group variation). Perhaps the only variable for which *d* may be asserted to be infinite—providing one accepts a certain physiological definition of maleness versus femaleness[7]—is "Do you have a Y chromosome (1=yes, 0=no)?" This is simply tautological. A Mars-versus-Venus case of disjunctively different "essences" could perhaps also be represented with *d*-values around 4 or 6, since these are the number of standard deviations that would have to lie between the means of two normal distributions for there to be very little or extremely little overlap between them. In fact, *d*-values this high do not commonly appear in the literature on sex differences in physiological or behavioral characteristics.

One of the largest commonly observable sex differences, for example, is in male and female heights, for which *d* has been estimated to be about 2.6 (Eliot 2009).[8] Given that heights are approximately normally distributed (and assuming equal variances), Figure 1 gives a picture of roughly how much difference—and how much overlap—is implied between the distribution of women's heights (dashed line) and men's heights (solid line).

FIGURE 1

Cohen's d = +2.6
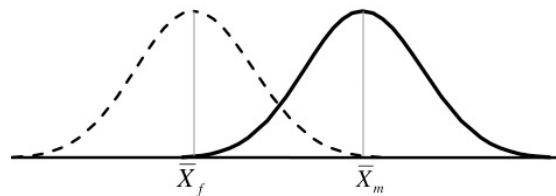
This is clearly a substantial difference—although not Mars-versus-Venus, since we not infrequently observe men and women who are the same height, at heights between the two means. The large *d*-value does, however, mean that it is relatively rare to observe men who are shorter than the average woman, or women taller than the average man. In cases when (strict) normality can be assumed, *d*-values can be easily converted into various other measures expressed as percentages of overlap, percentiles, ranks, correlations, or probabilities (Zakzanis 2001; Coe 2002). For example, in the above picture, 99.53% of the men's distribution lies above the female mean.

Suppose, instead, that d =.35. Then, again assuming normality and equal variances, the picture would be more like that shown in Figure 2.

---

[7] This definition is, however, disputed by those who identify as transsexual or genderqueer (Factor and Rothblum 2008).
[8] Throwing velocity is another characteristic associated with d>+2.0 (Hyde 2005). Presumably these estimates are based on data from the US or other industrialized societies.

FIGURE 2

Cohen's d= +.35

$\overline{X}_f \ \overline{X}_m$
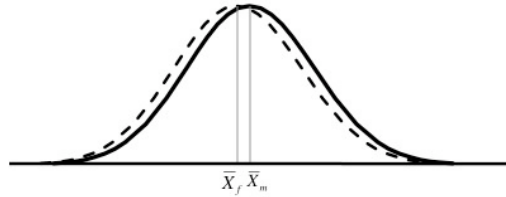
Clearly, this difference would be much less observable in everyday life. For example, in the above diagram, a considerably smaller share—64%--of the male distribution lies above the female mean.

Whether a given *d* value is "big," "moderate," or "small" depends a great deal on context and the purpose to which the interpretation of "difference" is being put. Suggested guidelines for qualitative interpretations are readily available in the literature, but should be approached with a great deal of caution. The value of *d*=+.35, for example, is clearly "small" in the sense that sex would be a quite unreliable signal of, say, *above average* ability in some skill being measured. To assume that a male advantage at the mean indicates that "men are more able" than women, when *d*=+.35, would be to ignore the 36% of men who are less able than the average woman, and the 36% of women who are more able than the average man. On the other hand, if being in the upper *tail* is the basis for employment promotions made on a tournament model, and there is a difference of this size in actual abilities—or merely in employers' perceptions of abilities, as in the case of discriminatory prejudices—*d* values in the range of low fractions could have a substantial impact (Martell, Lane et al. 1996).

The point here is to emphasize that while the language of simple difference tends, given the problems with generics discussed above, to tempt readers into Mars-versus-Venus thinking, *d*-values nuance the discussion of difference by offering one way of quantifying the degree of difference and reminding us of intra-group variability and hence ranges of overlap.

Note that *d* carries no implications about inference to a population. While it is mathematically derived from the same information as inferential statistics such as *t*- and *F*- statistics, a large *d*-value, considered on its own, contains no information relevant to making inferences. Like the *t*, its numerator is the difference between means, but unlike the *t*, its denominator is a (weighted) standard deviation, not a standard error. In diametrical contrast, the t statistic contains little information on its own about the substantive magnitude of the difference between the means (other than it is not exactly zero in the sample), since a large *t* may be in good part due to having a very large sample (and thus a small standard error).

Cohen's d, like any statistic, has some drawbacks and hazards in interpretation. It says nothing about differences in variance, skewness, or any other characteristics of the distributions. Its frequent pedagogical presentation in terms of normal distributions with

equal variances may lead to a temptation to infer additional characteristics (such as the degree of overlap) even when these conditions do not hold.

Note that *Cohen's d* can be computed—simply as a descriptive statistic—from distributions that do not look at all like the ones pictured above. For example, the d-value for the variable "Never bears a child, 1=true, 0=false" can be crudely estimated to be about +3.02 for adults in the United States.[9]

Index of Similarity

The *Index of Similarity (IS)* is an easily computable and understandable measure of overlap that does not rely on an assumption of normality. It can be calculated as

$$IS = Index\ of\ Similarity = 1 \ - \ \frac{1}{2}\left(\sum_i \left| \frac{f_i}{F} - \frac{m_i}{M} \right|\right)$$

where f_i/F is the proportion of females within category i, and m_i/M is the proportion of males in that same category. The categories may be qualitative (e.g. yes versus no answers), quantitative but limited in number (e.g., the number of lotteries entered out of nine offered), or might be continuous quantitative data aggregated into meaningful groups. *IS* has an intuitive interpretation as (in equal-sized groups) the proportion of the females and males that are similar, in the sense that their characteristics or behaviors (on this particular front) *exactly match up* with someone in the opposite sex group. If *IS*=.80, for example, it means that 80% of the women could be paired with a man with exactly the same behavior, or vice versa. If one imagines pairing up these matching subjects and setting them aside, it is clear that any differences in the overall distribution—and particularly, any difference in mean scores (when means can be calculated)—must be due to the behavior of remaining 20% of the subjects. *IS* is hence a direct measure of the *overlap* of the male and female distributions.

IS takes on values from 0 to 1. For the variable "do you have a Y chromosome," *IS*=0 for males and females if one assumes that chromosomes distinguish male from female. *IS*=1 for complete matching. *IS* is unlikely to be zero for non-definitional phenomena, even biologically related ones. For example, about 15% of US women are similar to men, in never bearing children.[10]

While one might expect *d* and *IS* to be inversely related—so that more "difference" corresponds to less "similarity"—this is not necessarily true, outside of the world of normal distributions with equal variances. For example, if a difference in means is due to a single large outlier, *d* could be substantial (that is, there is a large difference between the means) while *IS* exceeds .99 (most subjects are the same). On the other hand,

---

[9] According to U.S. Current Population Survey data from 2008, 17.8% of US women aged 40-44 never had a child (U.S. Census Bureau 2010). Since childbearing after age 40 is still relatively rare, one might guess, conservatively, that the overlap is in the area of around 15%. Thus, for men, the mean is 1 and standard deviation is zero, while for women the mean can be estimated at .15 with an implied standard deviation of .3582. Assuming equal sample sizes, d can be computed as +3.02.

[10] See previous footnote.

if the shapes or variances of the male and female distributions differ considerably, but in ways that have little overall effect on the means, the result could be a small value for *d* (little difference in means) and a small value for *IS* (but relatively few subjects "pair up"). When the underlying distributions are not normal with equal variance, the *d* and *IS* measures are complementary, giving two views into a complex reality.

Note that, unlike *d*, *IS* is non-directional. Knowing that, for example, 80% of the subjects of a study matched exactly does not yield any information about the direction of differences among the remaining 20%.

IS is derived from the "index of dissimilarity" (also called "Duncan's D") that has been long used to study racial housing segregation (Duncan and Duncan 1955). The same formula also underlies the "index of occupational segregation" used to study gender segregation of occupations (Reskin 1993; Blau, Ferber et al. 2010, 135). Mathematically, these are the part of the *IS* equation after the minus sign. They are commonly interpreted as the percent of either group (males or females; blacks or whites) who would have to *change* their zone of residence (for race) or occupation (for sex) for the responses to be identically distributed across the two groups. As these literatures have pointed out, one problem with such indices is that they are sensitive to the techniques and levels of aggregation used in defining categories (Reskin 1993, 243), so that care must be taken that these are not manipulated to create customized "results." The choice to define an index of *similarity* in the current essay, instead of dissimilarity, was based on the desire to create a countervailing symmetry with Cohen's *d*, which measures difference.

Magnitudes of Sex Difference and Similarity

A study was done of a number of published articles that deal with sex and risk, in order to answer the question, "*Given that* some studies have found statistically significant differences in measures of risk aversion or risk perception between groups of men and groups of women, what do these results imply about the *quantitative magnitudes* of the differences between means, and about the overlap of distributions?"

Study Design

The studies reviewed here were selected in the following manner. We started with those cited in Croson and Gneezy's (2009) meta-analysis, and then added much-cited older articles, did an EconLit search for newer articles, and added other articles as we encountered them in reference lists. The articles are primarily from the fields of economics and finance, though articles from decision science and psychology were also included if they have been cited in the economics literature and/or investigate similar phenomena as the economics literature. While it can be difficult to draw a clear line between "risk" and other behavioral phenomena such as competitiveness or sensation seeking, an effort was made to limit the analysis to studies that examined risk preferences (that is, degrees of willingness to take on risk) and/or risk perceptions (that is, variations in how hazardous a risk is perceived to be).

The point of this study being to supplement the usual reports about statistical significance with reports on substantive significance, we sought data from which we could compute *d* or *IS* values. In several cases, the necessary information for computing at least some of these statistics was present in the published papers.[11] In two cases we were able to get the necessary information from publically archived supplementary materials or datasets.[12] In addition, a number of authors generously shared their original data or specific statistics with us, on request.[13] In a number of cases, articles contained the results of multiple studies, for only some of which could the necessary information be obtained (either from the article or on request).

In other cases, however, and especially with older articles, we were unable to reach authors or authors told us that they could no longer access the relevant datasets. We necessarily, then, could not perform calculations at all for a number of articles—both those showing statistically significant gender differences in the usually expected direction (that is, of lesser male risk aversion or risk perception), as well as articles showing a lack of statistical significance, or statistically significant gender differences in the opposite of the usually expected direction (that is, greater male risk aversion or risk perception, e.g. Shubert, Brown et al. (1999)).[14] As time goes on, we continue to find articles that we would include in the analysis, were we to carry it further. In short, the present study should be interpreted as roughly representative of the literature, rather than as comprehensive. An accurate and complete meta-analysis may not, due to file drawer effects, as well as data availability problems, even be possible. Calculation on a representative group of articles is, however, sufficient for exploring what is meant by "difference."

The emphasis in our analysis is on differences between the "raw" distributions of men's and women's performance on various risk-related variables, before adjusting for covariates and, as much as possible, before dividing samples up into subsamples.[15]

---

[11] Articles that contained information sufficient to calculate these statistics (or, in some cases in the psychology literature, reported *d*-values directly) included Arano, Parker et al.(2010), Barsky, Juster et al. (1997), Bernasek and Shwiff (2001), Byrnes, Miller et al. (1999), Carr and Steele (2010), Eriksson and Simpson (2010), (2006), Harris, Jenkins et. al (2006), Lindquist and Säve-Söderbergh (2011), Meier-Pesti and Penz (2008), Olsen and Cox (2001), Powell and Ansic (1997), Rivers, Arvai et al. (2010), Ronay and Kim (2006), Sunden and Surette (1998), and Weaver, Vandello et al. (2012).

[12] We appreciate the standards for professional conduct and replication that lay behind the public availability of supplements to Eckel and Grossman (2008) and Holt and Laury (2002).

[13] We wish to express our appreciation to the authors of the following articles: Barber and Odean (2001), Beckmann and Menkhoff (2008), Booth and Nolen (2012), Borghans, Golsteyn et al. (2009), Dohmen, Falk et al. (2011) , Eriksson and Simpson (2010), Fehr-Duda, Gennaro et al. (2006), Finucane, Slovic et al. (2000), Gneezy, Leonard et al. (2009), Hartog, Ferrer-i-Carbonell et al. (2002), and Kahan, Braman et al. (2007).

[14] Additional studies we reviewed, but which did not result in statistics for Table 1, include Bruhin, Fehr-Duda et al (2010), Croson and Gneezy (2009), Flynn, Slovic et al. (1994), Levin, Snyder et al. (1988), Olofsson and Rashid (2011), Schubert, Brown et al. (1999), Sunden and Surette (1998), and Tanaka, Camerer et al. (2010).

[15] While regression coefficients may also be considered as measures of "effect size," these are not examined here for two reasons. The first is the practical consideration that information on raw distributions is available from, and comparable over, a wider range of published studies. The second reason is that regression results are potentially more influenced by invalid data mining (or "data dredging") practices.

Results for Cross-Sex Comparisons

Table 1 reports on an analysis of 24 published articles. The type of study done varied from analysis of survey questions asking people how they felt about various risks (including financial, environmental, and/or employment risks), to experimental studies in which subjects were offered lotteries of various types, to studies of financial asset allocations among risky or less-risky assets. The generally large sample sizes (often 200 into the tens of thousands) suggest that, if gender differences in mean scores are substantively large—or even, for very large samples, if the differences are substantively small—they will tend to appear as statistically significant.

TABLE 1
Magnitudes of Male vs. Female Differences and Similarities Related to Risk

Author(s)	Cohen's d	Index of Similarity	Type of Study	n (approx.)
Arano et al, 2010	NSS	—	Percent of actual retirement assets held in stock	400
Barber and Odean, 2001	-.09 to .26	—	Measures of riskiness of actual common stock holdings	38000
Barsky, Juster, et al., 1997	—	.98	Hypothetical question about taking a riskier job	12000
Beckmann and Menkhoff, 2008	NSS to +.46	.67 to .91	Survey questions about investment decisions	200
Bernasek and Shwiff, 2001	NSS	.87	Percent of actual retirement assets held in stock, and self-report	300
Booth and Nolen, 2012	NSS to .38	.84	Hypothetical investment, lottery	300
Borghans, Golsteyn, et al., 2009	.32 to .55	—	Lottery experiment	300
Byrnes, Miller, et al., 1999	-1.23 to +1.45 (Mean = +0.13)	—	Meta-analysis of 150 studies on risk in many domains	--
Dohmen, Falk, et al., 2011	NSS to .48	.80 to .88	General risk, lottery, and others (e.g., car driving, career)	500 to 22,000

Regression coefficients are not meaningful estimates of effect sizes if a researcher searches for and reports on only those combinations of variables or subsamples that yield statistically significant results, or the results that the researcher finds most plausible. Raw data is less influenced by these potential biasing factors. Note that this approach to reviewing the literature is quite different from the meta-regression analysis advocated by Stanley (2001), which focuses on the outcomes of regression specifications (rather than on raw data) and, in the example given in the article, on the sizes of interferential test statistics (rather than on substantive effect size).

Eckel and Grossman, 2008	.55 to 1.13	.60 to .80	Gambling questions	300
Eriksson and Simpson, 2010	.19 to .22	.89 to .91	Lottery experiment	200
Fehr-Duda, De Gennaro, et al., 2006	-.25 to NSS to .49	—	Lottery experiments	100
Finucane, Slovic, et al., 2000	.11 to .33	.86 to .93	Risk questions regarding hazards	1200
Harris, Jenkins, et al., 2006	-.34 to NSS to .74	—	Survey including health, recreational, gambling risks	700
Hartog, Ferrer-i-Carbonell, et al., 2002	.22 to .29	.85 to .96	Financial Lottery	1,500 to 13,000
Holt and Laury, 2002	NSS to .37	.83 to .86	Financial Lottery, hypothetical scenario	200
Kahan, Braman et al., 2007	.15 to .36	—	Perceptions of abortion and environmental risks	2000
Lindquist and Save-Soderbergh, 2011	NSS	—	Game show wagers	600
Meier-Pesti and Penz, 2008	NSS to .85	—	Hypothetical investments and gender risk question	150
Olsen and Cox, 2001	NSS to .65	.60 to .86	Survey of investment attitudes	200
Powell and Ansic, 1997	.06 to .17	.90 to .93	Insurance and market experiments, and survey	100
Rivers, Arvai et al, 2010	.25 to .31	—	Perceptions of health and environmental risks	400
Ronay and Kim, 2006	NSS to .44	—	Variety of questions and risk scenarios	50
Sunden and Surette, 1998	.08 to .16	.95 to .96	Survey on allocation of defined contribution assets	6000

Note: Adjusted so that positive d-values indicate relatively lesser risk-aversion (or risk perception) on the part of males, compared to females, on average. N's are approximate sample sizes of groups used for comparisons. NSS=No Statistically Significant difference. (See text for further explanation.)

Cohen's d, expressed such that a positive number signifies lesser mean male risk aversion or lesser male mean perception of risk, is reported in Table 1 for differences between means that were reported to be statistically significant (at a 10% level or better).[16] When the data allowed, these were computed for numeric variables and also for qualitative response variables (e.g. Likert scales) that the articles themselves treated as numeric and cardinal. Analysis was generally performed only on the subjects' own direct responses to survey questions, although in a few cases analysis was performed on the authors' univariate transformations of these variables.

Because many of the studies presented results for a number of different sampled groups and/or questions, the *d*-value results are presented as a range. The largest negative and positive (in absolute value) statistically significant numerical values are shown. "NSS" denotes that no statistically significant differences by sex were (also) found for some samples or variables, within a given study. The recording of NSS for an entire study indicates that, when evaluating the univariate responses for the sample as a whole, no statistically significant sex differences can be found. In these cases, the authors went on to find some evidence for differences in risk-taking in some subsamples and/or through multivariate analysis.

Note that a finding of a *d*-values exceeding +.50—that is, half a standard deviation, in favor of lesser male risk aversion—occurs in only six of 24 articles, and the finding of a difference of more than one standard deviation of difference occurs in only two. In most cases—and even within the same articles—smaller *d*-values are found. There are also many cases in which differences found are not statistically significant, and four articles in which differences that are statistically significant in the direction of greater female risk taking (*d*<0) are among the findings.

Table 1 also reports Indexes of Similarity for comparisons reported as statistically significant in the source articles. In some cases, *IS* values were computed for the same variables for which d-values were also computed, while in other cases they refer to different variables. Since these figures measure *similarity* but are only reported here for statistically significant *differences*, the numbers in Table 1 represent the low end of possible IS values that could be found in these data. *IS* values range from .60 to .98, with most studies yielding no values below .80. Because *IS* is non-directional, it is worth noting that one instance of IS=.67 (Beckmann and Menkhoff, 2008) is for a case where *fewer* men than women chose a risky option.

Figure 3 visually illustrates, as an example, data taken from one of the studies reviewed in Table 1. Beckmann and Menkhoff (2008) asked financial fund managers in four countries, "In respect of professional investment decisions, I mostly act…" giving

---

[16] A 10% level was chosen, rather than 5% or 1%, to give the existence of "difference" the maximum benefit of the doubt. Numeric values for *d* (or *IS*) were not calculated when differences were not statistically significant, because of the rather wild values that occured in some of the small samples. While in very large samples, one can assume that a lack of statistical significance is associated with a small *d-value*, in smaller samples, relatively large but highly unreliable d-values can occur, making reporting of their numerical values misleading.

them "six response categories ranging from 1=very risk averse to 6=little risk averse" (371). In the one country (Italy) for which the difference in mean response by sex to this question was statistically significant, calculation yields a *d* which is relatively substantial ($\approx$ .4) for this literature, and IS at the smaller end of the scale ($\approx$ .7). Many of the results in Table 1, therefore, represent cases of less "difference" or more "similarity" than that illustrated in Figure 3 (or even "difference" in the opposite direction).

FIGURE 3

$d \approx .4, \ IS \approx .7$
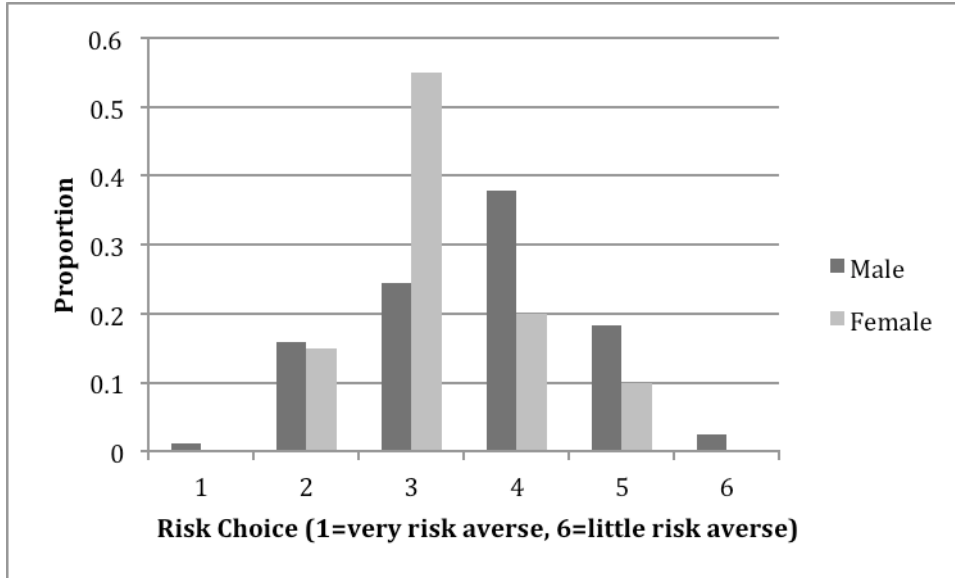
Are the indicators of difference and similarity in Table 1 small or large? While answering this could depend a great deal on a specific real-world context, the existence of "Men are from Mars, Women are from Venus" differences in risk-taking by sex can clearly be ruled out. To the extent that differences have been shown to exist in the literature—and the evidence is, as one can see, quite mixed—they are of a very much lesser order than, for example, the observable differences in the distributions of heights (d=+2.6) discussed earlier. Instead of difference, similarity seems to be the more prominent pattern, with well over half of men and women "matching up" on risk-related behaviors in every study. As one writer has quipped, perhaps "Men are from North Dakota, Women are From South Dakota" (Dindia 2006)—that is, men and women would seem to be more accurately regarded as being from neighboring states in the same country, with very much in common.

It should also be noted that nearly all the studies reviewed were based on men and women from Western industrialized societies.[17] In the cognitive science literature, doubts have been raised about the empirical validity of making generalizations about "human" behavior from such a WEIRD ("Western, Educated, Industrialized, Rich, and

---

[17] The exceptions in Table 1 are Beckmann and Menkoff (2008), who include a sample from Thailand; Eriksson and Simpson (2010), who include subjects from India; and possibly some studies reviewed in Byrnes et al (Byrnes, Miller et al. 1999).

Democratic" society) sample (Henrich, Heine et al. 2010). Further checking on behaviors presumed to be characteristic of males and females in cross-cultural context, before generalizing to all men and women, would seem to be warranted.

(Essential) Sex Difference, or a Result of Beliefs *about Difference?*

The generic explanation for observed differences in average risk-preference measures by sex (when they occur) is that greater risk aversion is a trait, characteristic, or essence shared by women by virtue of their being women. This, however, is not, as discussed above, the only possible explanation for differences in aggregate patterns by sex (when they occur). Differences that may appear at a cursory level to be due to sex "essences" may in fact be due (in part or completely) to a third, confounding variable, such as societal pressures to conform to gender expectations or to locations in a social hierarchy of power, or may no longer be seen when the sampling universe is broadened.

What if instead of looking empirically at the effect of *sex* on risk-taking (which may include both biological and cultural effects), one looks more explicitly at the effect of different socialization patterns or manipulations of *cultural gender* identifications or expectations? How do these compare? A literature in psychology has, in fact, grown up around this question.

One way to go about investigating this is to see if the degree of difference and sameness between the behaviors of men and women varies with cultural effects, either socially-generated or manipulated in a lab. Table 2 reports on the results of three such studies. Booth and Nolen (2012) studied the relationship between single-sex education versus co-education and the experimental lottery and investment behavior of girls and boys. While differences were observed between girls and boys educated in sex-integrated settings, on average, no statistically significant difference was found when both boys and girls received same-sex education.

TABLE 2

Magnitudes of Male vs. Female Differences and Similarities Related to Risk, With Confounding Cultural Effects

Author(s)	Cohen's d	Index of Similarity	Subgroup: Contrast	Study Type	n (approx.)
Booth and Nolen, 2012	NSS to .77	0.66 to .86	within co-educated: male vs. female	Hypothetical investment, lottery	150
Booth and Nolen, 2012	NSS	—	within single-sex educated: males vs. female	Hypothetical investment, lottery	100
Carr and Steele, 2010	1.3 to 1.7	—	within stereotype-threat: male vs. female	Experimental gambles	30

| Carr and Steele, 2010 | NSS | — | within stereotype-irrelevant: male vs. female | Experimental gambles | 30 |
| Gneezy. Leonard, et al., 2009 | NSS | NSS | within Maasai (Khasi) groups: male vs. female | Lottery experiment | 50 |

Note: Adjusted so that positive d-values indicate relatively lesser risk-aversion (or risk perception) on the part of the first group. Approximate N's are for each "within" subgroup. NSS=No Statistically --Significant difference.

Carr and Steele (2010) manipulated the gender framing of the experimental situation. They had male and female subjects experience either a "stereotype threat" situation or a "stereotype irrelevant" situation before measuring their risk-taking behavior using lottery games. In the "stereotype threat" situation, subjects were asked to record their gender before they were asked to play lottery games, which were described to them as testing their mathematical abilities. The extensive psychological literature on "stereotype threat" suggests that this may tend to erode women's performance, through causing women to worry about reinforcing a "women aren't good at math" stereotype. In the stereotype-irrelevant situation, subjects were not asked their gender until later, and the (same) experiment was described as being about puzzle-solving. Carr and Steele found very large differences (compared to Table 1—here *d* is as large as 1.7) when stereotype threat was activated, but no differences between men and women in risk-taking in the stereotype-irrelevant case.

While most of the studies in Table 1 were conducted on men and women from Western industrialized societies, Gneezy, Leonard, et al. (2009) studied subjects from a Maasai society in Africa and a Khasi society in South Asia. They found no statistically significant gender difference within either group, in a lottery experiment.[18] Evidence of the disappearance of "sex differences" upon the manipulation of cultural contexts makes the biological explanation appear less plausible, since an "essential" sex characteristic should presumably not vary with social context.

Another way of looking at the cultural gender issue is to ask about the degree to which differences *within* groups of males, or *within* groups of females, can be evoked through manipulating gender socialization, expectations, or identifications. The results of six such studies are summarized in Table 3.

---

[18] The major reported findings in their article are about competitiveness. On this variable, they found that women from the matrilineal Khasi society were *more* competitive than Khasi men, on average, while in the patrilineal Maasi society the pattern was reversed.

TABLE 3

Magnitudes of Differences of Males from Males, and Females from Females, Related to Risk, with Confounding Variables

Author(s)	Cohen's d	Index of Similarity	Subgroup: Contrast	Study Type	n (approx.)
Booth and Nolen, 2012	.31 to .71	.68 to .71	within females: single sex vs. coed	Hypothetical investment scenarios, lottery	150
Carr and Steele, 2010	.68 to 1.05	—	within female: no stereotype threat vs. threat	Experimental gambles	30
Kahan, Braman et al. 2007	.20 to .60	—	white males vs. everyone else	Perceptions of gun, abortion, and environmental risks	2000
Meier-Pesti and Penz, 2008	NSS to 0.91	—	within males: masculinity-primed vs. femininity-primed	Hypothetical investments and general risk question	30
Meier-Pesti and Penz, 2008	NSS	—	within females: masculinity-primed vs. femininity-primed	Hypothetical investments and general risk question	30
Ronay and Kim, 2006	.58 to 1.16	—	within males: with same-sex discussion vs. without	Variety of questions and risk scenarios	50
Ronay and Kim, 2006	NSS	—	within females: with same-sex discussion vs. without	Variety of questions and risk scenarios	50
Weaver, Vandello, et al., 2012	.57 to .74	—	within males: gender threat vs. gender affirmation	Gambling experiment	50

Note: Adjusted so that positive d-values indicate relatively lesser risk-aversion (or risk perception) on the part of the first group. Approximate N's are for each "within" subgroup. NSS=No Statistically Significant difference.

Booth and Nolen (2012) found statistically significant differences, on average, between girls who are single-sex educated and girls who are co-educated that tend to be larger than the differences they found between boys and girls taken as groups (reported in Table 1). Carr and Steele (2010) found differences, on average, between women in the stereotype-threat condition and women in the stereotype-irrelevant condition that in some cases exceed *d*=1.0.

Kahan, Braman et al. (2007) have investigated intersections of sex, race, and cultural worldviews following up on a finding that suggested that the most sizeable difference in risk perception tends to be, not between men and women, but between white males and everyone else (Finucane, Slovic et al. 2000). That is, non-white males' risk perceptions may be closer to that of females than to those of their fellow males. The *d*-values for white males versus everyone else in Kahan, Braman et al.'s (2007) study (Table 3) tend to be larger than the *d*-values they found for men versus women (Table 1). While not reviewed in this Table, an article by Flynn, Slovic et al. (1994) focuses more narrowly, and find the lowest perceptions of risk, on average, tend to be among white males who are also well educated, high income, politically conservative, and who trust authorities (1106). They suggest that risk attitudes may in this case be strongly influenced by relative positions in the social, political, and economic hierarchies: "Perhaps white males [on average] see less risk in the world because they create, manage, control, and benefit from so much of it" (1107).

Meier-Pesti and Penz (2008) primed subjects to think about gender by asking them to write either about a picture of a man in a business setting or a picture of a woman looking after a baby, while controls looked at a neutral picture. Subjects then completed a questionnaire about actual and hypothetical investment behavior and attitudes towards financial and other risk. While the gender-priming manipulation had little effect on female subjects, for some variables men who received masculinity priming revealed a statistically significantly higher propensity to take risk, on average, than those who were femininity-primed. The substantive magnitude, *d*=.91, is also quite a bit larger than the most of effects shown in Table 1.

Ronay and Kim (2006) found differences by sex, on average, on several exercises related to measuring risk attitudes and behavior (related to areas ranging from surfing to employment) that are in the range of others in the literature, as reported in Table 1. They also, however, had some subjects participate in a small group discussion with same-sex peers, while others answered the questions only individually. Males who participated in the all-male group discussion scored statistically (and substantively, in some cases, *d*>1.0) significantly higher on average, on risk–taking measures compared to both controls and their own pre-tests, though no effect of all-female group discussion was seen with females. The authors cite "social identity theory," which posits that "a desire to subscribe to [cultural norms of male daring] should be most pronounced when in the presence of one's own gender group" (Ronay and Kim 2006, 402), to suggest an explanation for these results.

The subjects of the study by Weaver, Vandello, et al. (2012) were all heterosexual males. Some were asked to test a feminine, scented hand lotion before doing a gambling experiment, creating a "gender threat" situation after which (some) men may feel a need to reestablish their masculinity.[19] Others were asked to test a power drill before doing the

---

[19] The assertion of bald statements and generalities based (invalidly) on aggregate analysis seems to be endemic to much of the literature, beyond economics. While it may be that only some men find hand lotion to be threatening, statements such as the following —perhaps unconsciously but still unfortunately— suggest that masculine identity is *universally* fragile: "Specifically, the apprehension that men feel about

experiment, creating a "gender affirmation" situation. The average amount bet and the average number of maximum bets was statistically significantly higher for the gender threat group compared to the gender affirmation group. Again, this within-sex substantive magnitude of difference ($d>.5$) is larger than many of the cross-sex effects seen in Table 1.

While sample sizes are relatively small and more replication is needed, taken together, the results shown in Tables 2 and 3 are strongly suggestive of sizeable effects of socialization and cultural beliefs about gender. These effects tend to exceed, in point estimates of quantitative magnitude, the sizes of the effects associated with sex difference per se (shown in Table 1). It may be doubted, then, whether the explanation of empirical sex differences (when they occur) requires hypotheses about metaphysical "essences." In most of the studies summarized in Table 1, no attention was paid to cultural or framing/priming effects. Determining the extent to which the differences found in such studies could be explained by differences in cross-cultural *beliefs* about gender, rather than sex per se, thus would require new research and/or replication with careful attention to these factors.

Digging for Difference?

As pointed out earlier, the fact that the hypothesis of different male and female essences is considered plausible (or even as definitively convincing) society-wide, can be expected to possibly lead to instances of confirmation bias. Combined with the fact that academic culture encourages the publication of only statistically significant differences, one may expect to see "differences" in the direction indicated by prior beliefs exaggerated in published studies.

This is, in fact, what our review found. The purpose of this essay is to point out profession-wide tendencies to diverge from the goal of objectivity, due in large part to commonly shared cognitive biases. The specific references in the discussion that follows, then, should be considered simply as examples of how a whole literature can drift in a particular direction due to widespread (though possibly erroneous) cultural beliefs, and not as accusations of failure on the part of individual authors.

Inaccurate citations of earlier literature

In reviewing the literature, one economics article states that "Previous surveys of economics…and psychology (James P. Byrnes, David C. Miller, and William D. Schafer 1999) report the same conclusions: women are more risk averse than men in the *vast majority* of environments and tasks" (Croson and Gneezy 2009, 449, emphasis added). Another article cites Byrnes, Miller et al. (1999) as demonstrating that "females' lower risk preferences and less risky behavior *is robust across a variety of contexts*" (Eriksson and Simpson 2010, 159, emphasis added). In fact, what Byrnes, Miller et al. (1999) actually concluded, after surveying studies of 322 different effects, was that "the majority

---

losing manhood status in other people's eyes leads them to compensate (or perhaps overcompensate) by taking greater risks and seeking immediate rewards" (Weaver, Vandello et al. 2012, 9).

(i.e., 60%) of the effects support the idea of greater risk taking on the part of males" and "a sizable minority (i.e., 40%) were either negative or close to zero" (Byrnes, Miller et al. 1999, 372).

Overemphasis on difference within a study's own results

In another study (Arano, Parker et al. 2010), a difference between married men and married women in the expected direction was highlighted in the text (Arano, Parker et al. 2010, 153), even though it was statistically insignificant. Meanwhile, differences between single men and single women went in the opposite direction. (That is, the point estimate of *d* was negative, though also not statistically significant). Regression analysis was then pursued on the married subsample and, with the addition of various covariates, a statistically significant regression coefficient on gender in the expected direction was found. No further investigation of the single subsample was reported.

In another study (Beckmann and Menkhoff 2008), the difference between male and female subjects on the most direct measure of risk aversion was statistically significant in only one of the four countries studied, and then only at a 10% significance level. Considering two other less direct questions as well, differences in only 5 out of 12 measures (four countries by three questions) were statistically significant (four at the 10% level and one at the 5% level). In a later section of the paper, it was found that females were statistically significantly *more* likely than males, on average, to *increase* investment risk taken on, under certain circumstances (378). Rather than taking this as evidence of possible higher male risk-aversion on average, a convoluted explanation was presented to justify this result as due to a greater presumed preference for conformity on the part of women ("strong ambition to stick close to the benchmark's performance" (Beckmann and Menkhoff 2008, 378)). While the data used in the study would seem to suggest that the evidence for greater female risk aversion is, at best, mixed, the article concludes that the data reveal "a victory for gender difference" and (367) "robust gender differences" (379) in the direction of women being "significantly more risk averse" (379).

Another study notes a "striking gender difference" in probability weights calculated from a combination of data and a particular theoretical framework, while skipping quickly over the fact that the distribution of men and women in major decision-making types was found to be "quite similar." (Bruhin, Fehr-Duda et al. 2010, 1402). While the discussion of the "striking gender difference" goes on for several pages, the practice of showing confidence bands—established in earlier sections of the article—is suddenly dropped. Earlier work on some of the same data (Fehr-Duda, Gennaro et al. 2006) had noted that male and female confidence bands for probability weights overlapped in 3.5 of the 4 treatments studied.

Another study reported, based on survey data, that "Women are significantly less willing to take risks than men in all domains." (Dohmen, Falk et al. 2011, 535) ($d \approx +.40$, $IS > .80$). The same study, however, also included a field experiment with hundreds of subjects. An analysis of that data reveals only a marginally statistically significant difference by sex on one risk question ($d = .17$, $IS = .88$, $p = .07$) and no statistically

significant sex difference (d=.05, *IS* = .90, p=.60) on the other. The published article does not report these weak to non-existent results.

In yet another study, findings of sex differences, in the aggregate, on one variable of interest were highlight throughout, while the lack of any statistically significant difference in a risk experiment was relegated to a footnote (Gneezy, Leonard et al. 2009, 1652).

Failure to consider confounding variables

While a number of studies mention the possibility that social factors may explain some of their results, these were usually not investigated. The experimental studies reviewed here, did not, in general, provide sufficient information about the experimental set-up for a reader to determine whether the sorts of cultural gender factors identified by Carr and Steele (2010)—e.g., asking about gender early on, or describing an exercise as mathematical—could have been important. Future articles could be improved by reporting these details.

In addition, considering how many of the studies deal with investment behavior, it is notable that few discuss the role of investment advice. A long tradition of treating "widows and orphans" (de Goede 2004) differently from male investors, as well as stereotypes about the presumed risk-aversion of female investors (Schubert, Brown et al. 1999, 385; Eckel and Grossman 2008, 15) may contribute to differences in average investment patterns, independent of investors' own inclinations.

Policy recommendations based on Mars-vs.-Venus stereotypes

Some studies base broad policy recommendations for women and men as generic categories on weak evidence about differences in aggregates. Two articles, for example, recommend special treatment for women investors and women investment advisers. In one, it was concluded that due to women's greater risk aversion, "female fund managers may be better suited to female customers who share their pattern in behavior than do men" (Beckmann and Menkhoff 2008, 381). This was in spite of finding mixed directions of the effect, no statistically significant difference larger than d=.5, and many cases of weak-to-no statistical significance. The results from the single statistically significant case (country) in which the difference in the means of the risk variable was statistically significant, in this study, were shown in Figure 3 (above, $d \approx .4$, *IS* $\approx .7$). A simulation based on those numbers, however, reveals the following: If we assume that female clients have the same distribution of (presumed) risk preferences as female fund managers (measured on a 6-point scale), the chance of a randomly selected female client being matched on risk aversion with a randomly selected female manager is only 37.5%. Women do not appear to have, as a group, just one "pattern of behavior." If the randomly selected manager is, instead, male, the chance of a match is not much lower, at 25%. In fact, the chance of a very risk averse female (i.e., one who selected one of the two lowest-risk options) being matched with manager of same risk aversion measure would actually

be slightly higher if the manager were randomly taken from male pool than from the female pool (17% vs. 15%).

A second article also recommended pairing female investors and advisors, even though *IS* values on the responses to risk questions can be calculated as exceeding .90 (Olsen and Cox 2001). It would seem that following the authors' recommendations would be a far inferior way of assigning fund managers to clients than simply asking the individuals about their risk preferences.

Examination of a narrow range of risks

The variety of types of risk studied is also quite limited, with lottery, gambling, and investment scenarios dominating the economic analysis. To what extent do these measure attitudes towards "risk"?

Many authors seem to assume the existence of a general sex-identified risk-aversion utility parameter applicable to all contexts—in one case, for example, hypothesizing that risk tendencies observed in lottery choices could be extrapolated to preferences concerning marriage and the afterlife (Hartog, Ferrer-i-Carbonell et al. 2002, 16). Other studies examine somewhat broader phenomena such as driving behavior or perceptions of environmental hazards. The studies that claim that "women are more risk averse than men," however, do not in general include in consideration areas of life in which women on average take on elevated risks relative to men, for example in pregnancy and childbirth or in relation to domestic violence.

The primary focus on lottery-type scenarios tends to draw attention towards situations of Knightian "risk," in which both payoffs and probabilities are known. "Uncertainty" is often narrowly interpreted in the literature as describing a case in which probabilities are not known, though the payoffs still are. Situations concerning the true sort of uncertainty generated simply by the fact that human beings live in a complex world that generates an unknown future, receive less attention. Yet unforeseen events— e.g., new inventions, bursting asset bubbles, or negative environmental consequences— regularly surprise us, and can be of very large economic consequence (Randall 2009; Taleb 2010). It may be argued that lottery experiments have the advantage of being more amendable to study, but if a focus on tractability drives economists to only "look under the lamppost" in studying risks, any generalization to larger-scale real-world concerns should be considered epistemologically suspect.

Policy Implications

If the economics literature has encouraged—or at least permitted the continuation of—exaggerated beliefs in the existence of greater risk-aversion among women among the public and among policy makers, what difference does this make? One case has already been mentioned: women may end up with inappropriate investment advice, or women fund-managers may end up with an unnecessarily restricted base of clients, if

their sex—rather than their actual preference—is taken as a signal of their investment priorities. A perception of greater risk aversion on the part of women may also lead to labor market discrimination, if women are assumed to be uninterested in occupations that involve high levels of risk-taking (Schubert, Brown et al. 1999). The ability of women to combat discrimination may also be compromised if labor market outcomes come to be explained away as a natural consequence of women's (inaccurately presumed) relative timidity, when discrimination is the actual cause.

The policy implications are wider than this, however. In the economics literature, having a relatively higher degree of risk aversion is nearly universally interpreted as a defect. Greater risk aversion is associated with an inability to "rationally" play lottery experiments (in an expected utility sense); with inadequate retirement portfolios (Bernasek and Shwiff 2001; Arano, Parker et al. 2010, 147), with neuroticism and a lack of ambition (Borghans, Golsteyn et al. 2009, 655); with an inability to advance in employment or entrepreneurship (Hartog, Ferrer-i-Carbonell et al. 2002, 24; Lindquist and Säve-Söderbergh 2011, 158; Booth and Nolen 2012, F56); and with a general tendency to let outdated, evolved habits get in the way of functioning well in modern societies (Eckel and Grossman 2002, 291).[20] The implication seems to be that men set the norm on risk behavior, to which women would do well to aspire.

On the other hand, the psychology, literature tends to examine both behaviors is which risk-taking has positive consequences and in which it has undesirable or dangerous consequences, such as reckless driving, smoking, or risky sexual behavior (Byrnes, Miller et al. 1999). This literature notes that risk-taking may "either be adaptive or maladaptive" (Byrnes, Miller et al. 1999, 368) and that risk-takers may be seen as either "heroes or fools" (Ronay and Kim 2006, 397). Too little risk aversion, it is noted, may be associated with "unrealistic illusions of control" that "suppress the feelings of anxiety that might otherwise serve to warn of danger" (Ronay and Kim 2006, 413).

An arguably even greater social harm, then, may come from reinforcing the stereotype of men as relatively more risk seeking. To the extent it is believed that "real men" should ignore risks and charge ahead, and cultures in the top leadership levels of business, finance, and government remain disproportionately masculine, decisions that lead to socially-excessive levels of risk may result (Ronay and Kim 2006, 415).

Could the financial crisis that began in 2008 be attributed, at least in part, to issues of sex and gender? In the wake of the crisis, several commentators asked whether women leaders would have prevented it, or whether it would have happened "if Lehman Brothers had been Lehman Sisters" (Kristof 2009; Morris 2009; Lagarde 2010). The evidence reviewed in this essay suggests, however, that the biological sex of the financial decision-makers or regulators is likely not the most important factor. Rather the one-sidedly masculine culture of much of the financial world, in which displays of prowess in stereotypically masculine-associated activities such as aggression and risk-taking are the

---

[20] A notable exception to this tendency is Barber and Odean (2001) which looks at a downside of "male" behavior—in this case an empirical tendency of men to trade stock more often, on average, than women, and thus reduce their returns. The authors interpret this as a sex-based difference in "overconfidence."

norm, may be more at issue. Such a macho culture has been well-documented in terms of masculine sexual imagery and sexual harassment (Chung 2010; McDowell 2010), and also, more recently, displayed in the increasing prominence of stereotypically masculine-associated technocratic and mathematical activities, such as modeling the values of financial derivatives (de Goede 2004, 207).[21] Since social identity theory predicts that exaggerated masculine-associated behavior may be most pronounced in all-male groups, greater prominence of women leaders in fields such as finance could play an important role in stabilizing the economy, for reasons quite other than any presumed innate risk-aversion. *Were* Wall Street firms and regulatory agencies such that they welcomed women and men as equal participants, this might indicate that societal gender stereotypes were breaking down. It might also be likely, then, that certain valuable characteristics and behaviors commonly stereotyped as feminine—such as carefulness—would be encouraged industry-wide, and certain inappropriate, reckless behaviors currently practices would be frowned upon, to the benefit of the industry and society.

An additional area in which greater risk-aversion could arguably be socially beneficial is the area of environmental policy, particularly as regards climate change. By assuming high discount rates, continuing steady GPD growth (by simple extrapolation), and infinite substitutability between resources, economists such as Nordhaus (2008), and Tol (2009) have painted a picture of a relatively controllable, rich future, even in the presence of climate change, and thus see rather little present need for ethical concern or mitigation action. Their intellectual framework has been adopted in recent U.S. governmental discussions of the social cost of carbon (U.S. Department of Energy 2010). Sunstein (2002-2003; 2005), the Administrator of the White House Office of Information and Regulatory Affairs, has written very dismissively about the Precautionary Principle advocated by many environmentalists (Randall 2009), advocating only very modest action on climate change. This is in sharp contrast to the emphatic advocacy coming from many scientists (and some economists) that swift and dramatic action is necessary to prevent further environmental and social damage (American Association for the Advancement of Science 2009; Union of Concerned Scientists 2010). Could it be that the continuing close adherence between mainstream economics models and methods and cultural norms of masculinity (Nelson 1992) has served to create "unrealistic illusions of control" (Ronay and Kim 2006, 413; DeMartino 2011, 193; Nelson in press)? As in the financial crisis case, the empirical work reviewed in this essay suggests that it would not be so much the presence of women in policy-making circles that would, through some innate tendency towards anxiety, lead to different sort of policies. Rather, a more accepting attitude towards women in leadership might also signal a breakdown of gender stereotypes that could cause a re-valuation of appropriate precaution and carefulness on the part of all decision-makers.

---

[21] Mathematical abilities is another area in which generic statements such as "Men are more able in math" have created an exaggerated public perception of "difference." Else-Quest, Hyde et al. (2010), in an international meta-analysis, find a range of effect sizes for math achievement of -0.42 to 0.40, with a mean *d*<0.15.

Conclusion

The statement "women are more risk averse than men" is fundamentally a metaphysical assertion about unobservable essences or characteristics, and therefore cannot be empirically proven or disproven. A review of the empirical literature, with attention paid to the misleading nature of generic beliefs and statements, the proper interpretation of statistical results, and the quantitative magnitudes of detectable differences and similarities, sheds doubt on whether statements such as these should have any place in an empirical science that aspires to objectivity. The widespread acceptance of such statements appears to perhaps be rooted more in confirmation bias than in reality.

In regards to future empirical work, the present essay suggests that more attention to the quantitative sizes of differences and similarities, and a more careful interpretation of aggregate results, could improve economists toolbox, and suggests the expanded use of two such mathematical tools. In regards to issues of risk, it is argued that exaggerated and stereotyped beliefs in the existence of sex-based differences may lead to suboptimal results in economic efficiency and equity. These may arise both through discriminatory treatment and through the encouragement of excessive risk-taking in important economic domains such as finance and the environment.

Julie A. Nelson is a Senior Research Fellow at the Global Development and Environment Institute, Tufts University, and a Professor of Economics at the University of Massachusetts, Boston. She received her Ph.D. in Economics from the University of Wisconsin. Inquiries can be directed to Julie.Nelson@tufts.edu.

REFERENCES

Acker, Joan (1990). "Hierarchies, Jobs, Bodies: A Theory of Gendered Organizations." *Gender and Society* 4(2): 139-158.

American Association for the Advancement of Science (2009). "Letter sent to members of the U.S. Senate by AAAS and 17 other U.S. scientific organizations.".

Arano, Kathleen, Carl Parker, et al. (2010). "Gender-Based Risk Aversion and Retirement Asset Allocation." *Economic Inquiry* 48(1): 147-155.

Archer, John (1996). "Comparing Women and Men: What is Being Compared and Why?" *American Psychologist* 51(2): 153-154.

Bakan, David (1955). "The general and the aggregate: A methodological distinction." *Perceptual and Motor Skills* 5: 211-212.

Bakan, David (1966). "The Test of Significance in Psychological Research." *Psychological Bulletin* 66(6): 423-437.

Barber, Brad M. and Terrence Odean (2001). "Boys will be Boys: Gender, Overconfidence, and Common Stock Investment." *The Quarterly Journal of Economics* 116(1): 261-292.

Barnett, Rosalind and Caryl Rivers (2004). *Same Difference: How Gender Myths Are Hurting Our Relationships, Our Children, and Our Jobs*. NY, Basic Books.

Barsky, Robert B., F. Thomas Juster, et al. (1997). "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study." *The Quarterly Journal of Economics* 112(2): 537-579.

Beckmann, Daniela and Lukas Menkhoff (2008). "Will Women Be Women? Analyzing the Gender Difference among Financial Experts." *Kyklos* 61(3): 364-384.

Bernasek, Alexandra and Stephanie Shwiff (2001). "Gender, Risk, and Retirement." *Journal of Economic Issues* 35(2): 345-356.

Blau, Francine D., Marianne A. Ferber, et al. (2010). *The Economics of Women, Men and Work*. Upper Saddle River, NJ, Pearson Prentice Hall.

Booth, Alison L. and Patrick Nolen (2012). "Gender Differences in Risk Behaviour: Does Nurture Matter?" *The Economic Journal* 122(558).

Borghans, Lex, Bart H. H. Golsteyn, et al. (2009). "Gender Differences in Risk Aversion and Ambiguity Aversion." *Journal of the European Economic Association* 7(2-3): 649-658.

Bruhin, Adrian, Helga Fehr-Duda, et al. (2010). "Risk and Rationality: Uncovering Heterogeneity in Probability Distortion." *Econometrica* 78(4): 1375-1412.

Byrnes, James P., David C. Miller, et al. (1999). "Gender Differences in Risk Taking: A Meta-Analysis." *Psychological Bulletin* 125(3): 367-383.

Carr, Priyanka B. and Claude M. Steele (2010). "Stereotype Threat Affects Financial Decision Making." *Psychological Science* 21: 1411-1416.

Chung, Christine Sgarlata (2010). "From Lily Bart to the Boom Boom Room: How Wall Street's Social and Cultural Response to Women Has Shaped Securities Regulation." *Harvard Journal of Law & Gender* 33(1): 175-245.

Coe, Robert (2002). It's the Effect Size, Stupid: What effect size is and why it is important. *Annual Conference of the British Educational Research Association*. University of Exeter, England.

Croson, Rachel and Uri Gneezy (2009). "Gender Differences in Preferences." *Journal of Economic Literature* 47(2): 448-474.

Cross, Catherine P., Lee T. Copping, et al. (2011). "Sex Differences in Impulsivity: A Meta-Analysis." *Psychological Bulletin* 137(1): 97-130.

de Goede, Marieke (2004). "Repoliticizing financial risk." *Economy and Society* 33(2): 197-217.

Deaton, Angus S. (2009). Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development. *NBER Working Paper Series*. Cambridge, MA, NBER.

DeMartino, George F. (2011). *The Economist's Oath: On the Need for and Content of Professional Economic Ethics*. Oxford, Oxford University Press.

Dindia, Kathryn (2006). Men are from North Dakota, Women are from South Dakota. *Sex Differences and Similarities in Communication*. K. Dindia and D. J. Canary. New York, Routledge**:** 3-18.

Dohmen, Thomas, Armin Falk, et al. (2011). "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences." *Journal of the European Economic Association* 9(3): 522-550.

Duncan, Otis D. and Beverly Duncan (1955). "A Methodological Analysis of Segregation Indexes." *American Sociological Review* 20(2): 210-217.

Durlak, Joseph A. (2009). "How to Select, Calculate, and Interpret Effect Sizes." *Journal of Pediatric Psychology* 34(9): 917-928.

Eagly, Alice H. (1995). "The Science and Politics of Comparing Women and Men." *American Psychologist* 50(3).

Eckel, Catherine C. and Philip J. Grossman (2002). "Sex differences and statistical stereotyping in attitudes toward financial risk." *Evolution and Human Behavior* 23: 281-295.

Eckel, Catherine C. and Philip J. Grossman (2008). "Forecasting risk attitudes: An experimental study using actual and forecast gamble choices." *Journal of Economic Behavior & Organization* 68(1): 1-17.

Eliot, Lise (2009). *Pink Brain, Blue Brain: How Small Differences Grow into Troublesome Gaps--And What We Can Do About It*. Boston, Houghton Mifflin Harcourt.

Else-Quest, Nicole M., Janet Shibley Hyde, et al. (2010). "Cross-National Patterns of Gender Differences in Mathematics: A Meta-Analysis." *Psychological Bulletin* 136(1): 103-127.

Eriksson, Kimmo and Brent Simpson (2010). "Emotional reactions to losing explain gender differences in entering a risky lottery." *Judgment and Decision Making* 5(3): 159-163.

Factor, Rhonda J. and Esther Rothblum (2008). "Exploring Gender Identity and Community Among Three Groups of Transgender Individuals in the United States: MTFs, FTMs, and genderqueers." *Health Sociology Review* 17(Community, Family, Citizenship and the Health of LGBTIQ People): 235-253.

Fehr-Duda, Helga, Manuele De Gennaro, et al. (2006). "Gender, Financial Risk, and Probability Weights." *Theory and Decision* 60: 283-313.

Fine, Cordelia (2010). *Delusions of Gender: How Our Minds, Society, and Neurosexism Create Difference*. NY, W.W. Norton.

Finucane, Melissa L., Paul Slovic, et al. (2000). "Gender, race, and perceived risk: The 'white male' effect." *Health, Risk & Society* 2(2): 159-172.

Flynn, James, Paul Slovic, et al. (1994). "Gender, race, and perception of environmental health risks." *Risk Analysis* 14(6): 1101-1108.

Gelman, Susan A. (2005). "Essentialism in Everyday Thought." *Psychological Science Agenda*(May): 1-6.

Gneezy, Uri, Kenneth L. Leonard, et al. (2009). "Gender Differences in Competition: Evidence from a matrilineal and patriarchal society." *Econometrica* 77(5): 1637-1664.

Gray, John (1993). *Men are from Mars, Women are from Venus* NY, HarperCollins.

Harris, Christine R., Michael Jenkins, et al. (2006). "Gender Differences in Risk Assessment: Why do Women Take Fewer Risks than Men?" *Judgment and Decision Making* 1(1): 48-63.

Hartog, Joop, Ada Ferrer-i-Carbonell, et al. (2002). "Linking Measured Risk Aversion to Individual Characteristics." *Kyklos* 55(1): 3-26.

Haslanger, Sally (2011). Ideology, Generics, and Common Ground. *Feminist Metaphysics: Explorations in the Ontology of Sex, Gender and the Self* C. Witt. Dordrecht, Springer**:** 179-208.

Henrich, Joseph, Steven J. Heine, et al. (2010). "The weirdest people in the world?" *Behavioral and Brain Sciences* 33(2/3): 1-23.

Holt, Charles A. and Susan K. Laury (2002). "Risk Aversion and Incentive Effects." *The American Economic Review* 92(5): 1644-1655.

Hyde, Janet Shibley (2005). "The Gender Similarities Hypothesis." *American Psychologist* 60(6): 581-592.

Hyde, Janet Shibley and Elizabeth Ashby Plant (1995). "Magnitude of Psychological Gender Differences: Another Side to the Story." *American Psychologist* 50(3): 159-161.

Joop Hartog, Ada Ferrer-i-Carbonell and Nicole Jonker (2002). "Linking Measured Risk Aversion to Individual Characteristics." *Kyklos* 55(1): 3-26.

Kahan, Dan M., Donald Braman, et al. (2007). "Culture and Identity-Protective Cognition: Explaining the White-Male Effect in Risk Perception." *Journal of Empirical Legal Studies* 4(3): 465-505.

Khemlani, Sangeet, Sarah-Jane Leslie, et al. (2009). *Generics, prevalence, and default inferences*. CogSci 2009: 31st Annual Meeting of the Cognitive Science Society, Amsterdam.

Khemlani, Sangeet, Sarah-Jane Leslie, et al. (2012). "Inferences about members of kinds: The generics hypothesis." *Language and Cognitive Processes* 27(6): 887-900.

Kimmel, Michael S. (2000). *The Gendered Society*, Oxford University Press.

Kristof, Nicholas D. (2009). Mistresses of the Universe. *New York Times*.

Lagarde, Christine (2010). "Women, Power and the Challenge of the Financial Crisis." *International Herald Tribune*: Op-Ed.

Leslie, Sarah-Jane (2008). "Generics: Cognition and Acquisition." *Philosophical Review* 117(1): 1-47.

Leslie, Sarah-Jane (forthcoming). "The Original Sin of Cognition: Fear, Prejudice and Generalization." *The Journal of Philosophy*.

Levin, Irwin P., Mary A. Snyder, et al. (1988). "The Interaction of Experiential and Situational Factors and Gender in a Simulated Risky Decision-Making Task." *The Journal of Psychology and Financial Markets* 122(2): 173-181.

Lindquist, Gabriella Sjögren and Jenny Säve-Söderbergh (2011). ""Girls will be Girls", especially among Boys: Risk-taking in the "Daily Double" on Jeopardy." *Economics Letters* 112(2): 158-160.

Martell, Richard F., David M. Lane, et al. (1996). "Male-Female Differences: A Computer Simulation." *American Psychologist* 51(2): 157-158.

McDowell, Linda (2010). "Capital Culture Revisited: Sex, Testosterone, and the City." *International Journal of Urban and Regional Research* 34(3): 652-8.

Meier-Pesti, Katja and Elfriede Penz (2008). "Sex or Gender? Expanding the sex-based view by introducing masculinity and femininity as predictors of financial risk taking." *Journal of Economic Psychology* 29: 180-196.

Miller, Jane E. and Yana van der Meulen Rodgers (2008). "Economic Importance and Statistical Significance: Guidelines for Communicating Empirical Research." *Feminist Economics* 14(2): 117-49.

Morris, Nigel (2009). Harriet Harman: 'If only it had been Lehman Sisters'. *The Independent*. London.

Nelson, J. A. (in press). "Ethics and the economist: What climate change demands of us." *Ecological Economics* Working paper available at http://ase.tufts.edu/gdae/pubs/wp/11-02ethicsandeconomists.pdf.

Nelson, Julie A. (1992). "Gender, Metaphor, and the Definition of Economics." *Economics and Philosophy* 8: 103-125.

Nordhaus, William D. (2008). *A Question of Balance: Weighing the Options on Global Warming Policies*. New Haven, Yale University Press.

Olofsson, Ann and Saman Rashid (2011). "The White (Male) Effect and Risk Perception: Can Equality Make a Difference?" *Risk Analysis* 31(6): 1016-1032.

Olsen, Robert A. and Constance M. Cox (2001). "The Influence of Gender on the Perception and Response to Investment Risk: The Case of Professional Investors." *The Journal of Psychology and Financial Markets* 2(1): 29-36.

Powell, Melanie and David Ansic (1997). "Gender differences in risk behaviour in financial decision-making: An experimental analysis." *Journal of Economic Psychology* 18: 605-628.

Randall, Alan (2009). "We Already Have Risk Management--Do We Really Need the Precautionary Principle?" *International Review of Environmental and Resource Economics* 3(1): 39-74.

Reskin, Barbara (1993). "Sex Segregation in the Workplace." *Annual Review of Sociology* 19: 241-270.

Rivers, Louie, Joseph Arvai, et al. (2010). "Beyond a Simple Case of Black and White: Searching for the White Male Effect in the African-American Community." *Risk Analysis* 30(1): 65-77.

Ronay, Richard and Do-Yeong Kim (2006). "Gender differences in explicit and implicit risk attitudes: A socially facilitated phenomenon." *British JOurnal of Social Psychology* 45(2): 397-419.

Schubert, Renate, Martin Brown, et al. (1999). "Financial Decision-Making: Are Women Really More Risk-Averse?" *American Economic Review* 89(2): 381-385.

Stanley, T. D. (2001). "Wheat From Chaff: Meta-Analysis As Quantitative Literature Review." *2001* 15(3): 131-150.

Sunden, Annika E. and Brian J. Surette (1998). "Gender Differences in the Allocation of Assets in Retirement Savings Plans." *The American Economic Review* 88(2): 207-211.

Sunstein, Cass R. (2002-2003). "The Paralyzing Principle." *Regulation* Winter

Sunstein, Cass R. (2005). *Laws of Fear: Beyond The Precautionary Principle*. NY, Cambridge University Press.

Taleb, Nassim Nicholas (2010). *The Black Swan: The Impact of the Highly Improbable*. NY, Random House.

Tanaka, Tomomi, Colin F. Camerer, et al. (2010). "Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam." *American Economic Review* 100(1): 557-571.

Tol, Richard S. J. (2009). "The Economic Effects of Climate Change." *Journal of Economic Perspectives* 23(2): 29-51.

U.S. Census Bureau (2010). Current Population Survey - June 2008, Detailed Tables. P. Division.

U.S. Department of Energy (2010). "Final Rule Technical Support Document (TSD): Energy Efficiency Program for Commercial and Industrial Equipment: Small Electric Motors," Appendix 15A (by the Interagency Working Group on Social Cost of Carbon): "Social Cost of Carbon for Regulatory Impact Analysis Under Executive Order 12866".

Union of Concerned Scientists (2010). U.S. Scientists and Economists' Call for Swift and Deep Cuts in Greenhouse Gas Emissions.

Weaver, Jonathan R., Joseph A. Vandello, et al. (2012). "Intrepid, Imprudent, or Impetuous?: The Effects of Gender Threats on Men's Financial Decisions." *Psychology of Men & Masculinity* Advance online publication.: 1-12.

Wilkinson, Leland and Task Force on Statistical Inference (1999). "Statistical Methods in Psychology Journals: Guidelines and Explanations." *American Psychologist* 54(8): 594-604 (1-24 in downloadable document).

Zakzanis, Konstantine K. (2001). "Statistics to tell the truth, the whole truth, and nothing but the truth: formulae, illustrative numerical examples, and heuristic interpretation of effect size analyses for neuropsychological researchers." *Archives of Clinical Neuropsychology* 16: 653-667.

Ziliak, Stephen T. and Deirdre N. McCloskey (2004). "Size matters: the standard error of regressions in the American Economic Review." *The Journal of Socio-Economics* 33: 527-546.

Recent Papers in this Series: