

University of Massachusetts Boston

From the Selected Works of Julie A. Nelson

August, 2013

Are Women Really More Risk-Averse than Men? A Re-Analysis of the Literature Using Expanded Methods

Julie A. Nelson, *University of Massachusetts Boston*



Available at: https://works.bepress.com/julie_nelson1/26/

Are Women Really More Risk-Averse than Men?

A Re-Analysis of the Literature Using Expanded Methods

Julie A. Nelson
Professor and Chair
Department of Economics
University of Massachusetts Boston
100 Morrissey Blvd.
Boston, MA 02125 USA
Julie.nelson@umb.edu
(617)287-6925 phone
(617)287-6976 fax

August 6, 2013

Abstract

While a substantial literature in economics and finance has concluded that “women are more risk averse than men,” this conclusion merits investigation. After briefly clarifying the difference between making generalizations about groups, on the one hand, and making valid inferences from samples, on the other, this essay suggests improvements to how economists communicate our research results.

Supplementing findings of statistical significance with quantitative measures of both substantive difference (*Cohen's d*, a measure in common use in non-Economics literatures) and of substantive overlap (the *Index of Similarity*, newly proposed here) adds important nuance to the discussion of sex differences. These measures are computed from the data on men, women, and risk used in 24 published articles from economics, finance, and decision science. The results are considerably more mixed and overlapping than would commonly be inferred from the broad claims made in the literature, with standardized differences in means mostly amounting to considerably less than one standard deviation, and the degree of overlap between male and female distributions generally exceeding 80%. In addition, studies that look at contextual influences suggest that these contribute importantly to observations of differences both between and within the sexes.

Keywords: sex, gender, difference, similarity, risk, effect size

JEL codes: C9, D81, J16, B41

1. The Issue

Many studies in economics and finance have concluded that "women are more risk averse than men," generally based on findings of a statistically significant difference between men's and women's behavior, on average, in lottery and gambling experiments, and/or studies of investment behavior. Croson and Gneezy (2009), for example, affirm this conclusion in their survey of experimental studies, in which they report for each study the direction of the statistically significant differences found, along some details of the design.¹ From this evidence it is now often taken as a truism that women, by virtue of their sex, are categorically and importantly different from men in the trait of risk preference (e.g., discussions in Olen 2012; Nelson 2013).

Such a conclusion could be misleading, for a number of reasons. First, statistical significance may be confused with substantive significance. Differences between men and women are often popularly understood in the dichotomous way popularized by such books as John Gray's *Men are From Mars, Women are From Venus* (1993) or Simon Baron-Cohen's *The Essential Difference* (Baron-Cohen 2003), which portray men and women as having markedly distinct natures. Yet information on statistical significance alone says nothing about the actual *degree* of difference observed. Second, observed differences, on average, between men and women that are believed to be caused by biological sex may in fact be, all or in part, the result of other confounding variables, including the upbringing of a study's subjects and the cultural context and framing of the study itself. Third, publication bias may lead to studies that find statistically significant differences, on average, between the sexes

disproportionately appearing in journals, while studies that fail to find statistical significance are filed away.

This study examines at the first and second issues—the substantive magnitude of observed differences and the influence of confounding variables—through a brief discussion of the issues involved, followed by a re-analysis of 24 published articles from economics, finance, and decision science related to men, women, and risk. Supplementing findings of statistical significance with quantitative measures of both substantive difference (*Cohen's d*, a measure in common use in non-Economics literatures) and of substantive overlap (the *Index of Similarity*, newly proposed here) yields results that are considerably more mixed and overlapping than the "Mars-Versus-Venus" picture that is commonly inferred. In addition, examining both the across-sex and within-sex differences (or lack thereof) found in a set of studies in which variables related to cultural context are manipulated sheds some doubt on the conclusion that risk preferences are simply a sex-linked trait. Because of space constraints, the (third-mentioned) issue of publication bias, as well as more detail about the formation of generalizations and stereotypes, are discussed elsewhere (Nelson 2013).²

The emphasis in the current study is on differences between the "raw" distributions of men's and women's scores on various risk-related variables, before adjusting for covariates and, as much as possible, before dividing samples up into subsamples. While regression coefficients are also measures of substantive effects, 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. 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. This approach to reviewing the literature is hence different from formal meta-regression analysis (T. D. Stanley 2001; T.D. Stanley et al. 2013). For reasons of space, many details about study design in the articles reviewed (e.g., the exact design of lotteries) are also not reported or examined here. While these might be informative in explaining why results differ *among* studies, the overall focus of this essay is on the verifiability of the claim, arising from the studies *as a whole*, that “women are more risk averse than men.”

2. Generalization Versus Inference

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, they in fact can convey very different meanings.

Statement A is a narrow statement that can be factually correct within the confines of a particular study. “Men” and “women” in this statement are simply sets

of individuals (identified in the data by these labels), and the difference found is clearly stated as referring to group means.

Statement B, however, is a broad, generic statement that—according to research in linguistics, philosophy, and psychology—is often understood as implying stable characteristics of people according to their sex. "Women" in this second sense "refers to a category rather than a set of individuals," and members of such a generic category are assumed to share "essential qualities" (Gelman 2005). In the current example, the statement would seem to 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. A statement phrased as a generic and accepted as true also, research demonstrates, predisposes people to believe that individual members of a class will have the stated property (Khemlani, Leslie et al. 2009, 447).

Do researchers *intend* their "Women are more risk averse than men" conclusions to be understood in the much stronger, generic sense? Perhaps not. It could be that Statement B is, within the profession, simply understood as a convenient shorthand for Statement A.³ The fact that it may be widely misinterpreted by the public, however, especially in ways that may reinforce potentially deleterious stereotypes, should be an argument for more careful reporting. The techniques explored later in this essay are offered in the hope of improving the quality of economists' contributions to public understanding.

The same point may also be seen as illustrating the distinction between Fisherian statistical inference and inductive reasoning, as helpfully contrasted by

Bakan (1966). Fisherian inference means going from sample results concerning a aggregate, such as a sample difference in means, to inferences about the corresponding population aggregate. Statement A is an example of such Fisherian statistical inference, justifying (with a given level of confidence, and assuming unbiased reporting) making inferences about the *group means* in the population from which the samples were taken. To reason inductively, on the other hand, means to go from specific observations to hypothesizing general propositions that invite conclusions about the *nature* of the subjects of study. Statement B is such a statement about natures, including the presumed "nature" of every individual man and women. Statement A does not logically imply Statement B.

3 Statistical Tools for Investigating Sameness and Difference

While it is widely recognized that "substantive significance" and "statistical significance" are two different things, discussions of the size and importance of an estimated (statistically significant) sex difference in risk preferences are rare in the economics literature.⁴ 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 binary 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 sections, 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.

3.1 Cohen's d

Cohen's d is one measure of "effect size," expressing the magnitude of a difference between means (e.g., Byrnes et al. 1999; Cross et al. 2011; Wilkinson and Task Force on Statistical Inference 1999; Hyde 2005). 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

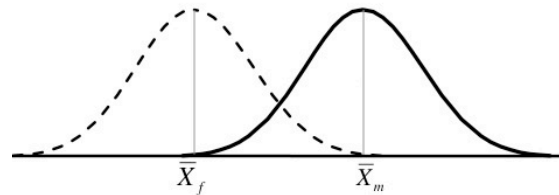
$$d = \frac{\bar{X}_m - \bar{X}_f}{s_p}$$

where \bar{X}_m is the male mean, \bar{X}_f is the female mean, and s_p is the pooled standard deviation, a measure of the average within-group variation.⁵ 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.

Cohen's d can, in theory, take on values from $-\infty$ to $+\infty$, 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 sex-related variable for which d may be asserted to be infinite—providing one accepts a certain physiological definition of maleness versus femaleness⁶—is "Do you have a Y chromosome (1=yes, 0=no)?" This is simply tautological. A "Men are from Mars, Women are from Venus" case of disjunctively different "essences" could perhaps also be represented with d -values of 4 or 6 or more, 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 mean sex differences in either 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).⁷ 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).

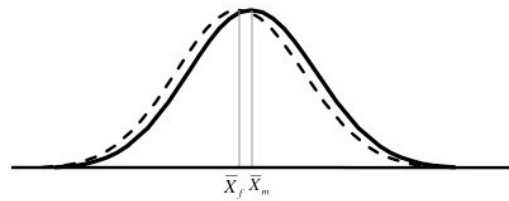
FIGURE 1
Cohen's $d = +2.6$



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.

FIGURE 2
Cohen's $d = +.35$



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 et al. 1996).

The point here is to emphasize that while the language of simple difference tends to tempt consumers of research into Mars-versus-Venus thinking, d -values

nuance the discussion of difference by offering one way of quantifying the degree of difference and reminding the reader 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 +3.0 for adults in the United States.⁸

3.2 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

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

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.⁹

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, if the shapes or variances of the male and female distributions differ considerably, but in ways which 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 et al. 2010). 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), 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.

4. Magnitudes of Sex Difference and Similarity

The current study re-analyzes 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?*"

4.1 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. Many of the

articles—and especially the articles that have appeared in highly-ranked journals such as *The Economic Journal* or the *Quarterly Journal of Economics*—use experimental methods. Because both experimental and non-experimental evidence tends to be cited in literature reviews, however, it was decided to include articles using all types of data in this study. 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.¹⁰ In two cases we were able to get the necessary information from publically archived supplementary materials or datasets.¹¹ In addition, a number of authors generously shared their original data or specific statistics with us, on request.¹² 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)).¹³ As time went on, we continued to find articles that we would include in the analysis, were we to carry it further.¹⁴ 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 the difficulty of getting unpublished studies 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."

4.2 Results for Cross-Sex Comparisons

Table 1 reports on an analysis of 24 published articles. Many were experimental studies in which subjects were offered lotteries of various types,. Others analyzed survey questions asking people how they felt about various risks (including financial, environmental, and/or employment risks), or studied 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.)
Byrnes, Miller, et al., 1999	-1.23 to +1.45 (Mean = +0.13)	—	Meta-analysis of 150 studies on risk in many domains	--
Harris, Jenkins, et al., 2006	-.34 to NSS to .74	—	Survey including health, recreational, gambling risks	700
Fehr-Duda, De Gennaro, et al., 2006	-.25 to NSS to .49	—	Lottery experiments	100
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
Arano et al, 2010	NSS	—	Percent of actual retirement assets held in stock	400
Bernasek and Shwiff, 2001	NSS	.87	Percent of actual retirement assets held in stock, and self-report	300
Lindquist and Save-Soderbergh, 2011	NSS	—	Game show wagers	600
Holt and Laury, 2002	NSS to .37	.83 to .86	Financial Lottery, hypothetical scenario	200
Booth and Nolen, 2012	NSS to .38	.84	Hypothetical investment, lottery	300
Ronay and Kim, 2006	NSS to .44	—	Variety of questions and risk scenarios	50
Beckmann and Menkhoff, 2008	NSS to .46	.67 to .91	Survey questions about investment decisions	200
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
Olsen and Cox, 2001	NSS to .65	.60 to .86	Survey of investment attitudes	200
Meier-Pesti and Penz, 2008	NSS to .85	—	Hypothetical investments and gender risk question	150
Powell and Ansic, 1997	.06 to .17	.90 to .93	Insurance and market experiments, and survey	100

Sunden and Surette, 1998	.08 to .16	.95 to .96	Survey on allocation of defined contribution assets	6000
Finucane, Slovic, et al., 2000	.11 to .33	.86 to .93	Risk questions regarding hazards	1200
Kahan, Braman et al., 2007	.15 to .36	—	Perceptions of abortion and environmental risks	2000
Eriksson and Simpson, 2010	.19 to .22	.89 to .91	Lottery experiment	200
Hartog, Ferrer-i-Carbonell, et al., 2002	.22 to .29	.85 to .96	Financial Lottery	1,500 to 13,000
Rivers, Arvai et al, 2010	.25 to .31	—	Perceptions of health and environmental risks	400
Borghans, Golsteyn, et al., 2009	.32 to .55	—	Lottery experiment	300
Eckel and Grossman, 2008	.55 to 1.13	.60 to .80	Gambling questions	300

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).¹⁵ 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 sometimes went on to find some evidence for differences in risk-taking in some subsamples and/or through multivariate analysis. The studies are listed in order according to the level of support they give for the proposition that males on average have a consistently stronger preference for risk, beginning with contradictory (i.e., including some evidence of greater mean *female* risk preference, $d < 0$), through a lack of evidence for a difference, to support.

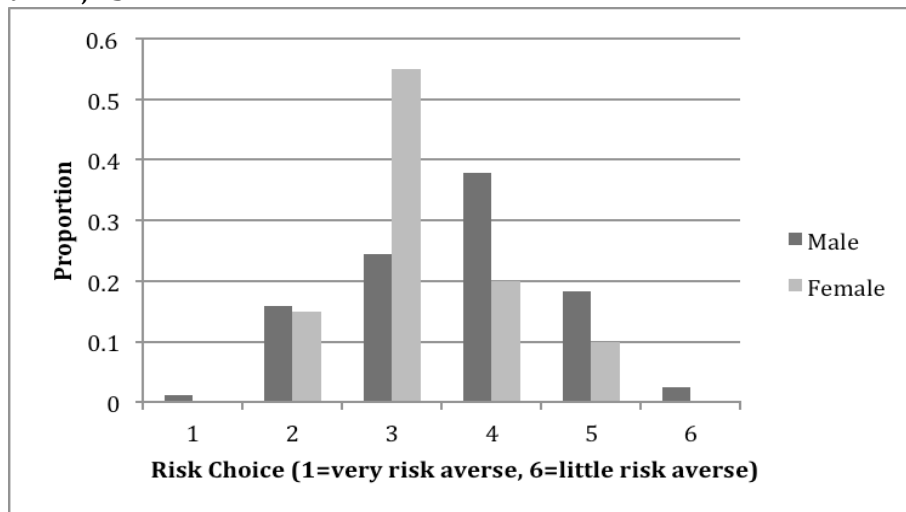
Note that a finding of a *d*-values exceeding $+0.50$ —that is, half a standard deviation, in favor of lesser male risk aversion—occurs in only six of 24 articles, and the finding of a difference of more than one standard deviation of difference occurs in only two. In the vast majority of cases—and even within the same articles that yield those relatively large *d*-values—smaller *d*-values and/or cases that lack statistical significance are found. In four articles differences that are statistically significant in the direction of greater female risk taking 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 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.¹⁶ 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 Democratic" society) sample (Henrich 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.

5. Empirical Studies on the Importance of Culture

The essence-based explanation proposed for observed differences in average risk-preference measures by sex (when they occur) is that greater risk aversion is a trait or characteristic shared by women by virtue of their being women. This, however, is not, as discussed above, the only possible reason why differences in aggregate patterns by sex might be observed. Differences that may appear at a cursory level to be due to sex "essences" may in fact be due (in part or completely) to some 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.

While the studies in Table 1 examine the effect of *sex* on risk-taking (which may include both biological and cultural effects), literatures exist that examine the possible effects of *cultural gender* identifications or expectations. Some of these generate different effects through experimental manipulations, while others look at the evidence cross-culturally.

5.1 Cross-Sex Studies of Cultural Effects

One way to go about investigating cultural effects is to see if the degree of difference and sameness between the behaviors of men and women varies with social context, 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. No statistically significant difference was found when both boys and girls received same-sex education, though differences were observed between girls and boys educated in sex-integrated settings, on average.

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	—	within single-sex educated: males vs. female	Hypothetical investment, lottery	100
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
Booth and Nolen, 2012	NSS to .77	0.66 to .86	within co-educated: male vs. female	Hypothetical investment, lottery	150
Carr and Steele, 2010	1.3 to 1.7	—	within stereotype-threat: male vs. female	Experimental gambles	30

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 no differences between men and women in risk-taking in the stereotype-irrelevant case, but very large differences (compared to Table 1—here d is as large as 1.7) when stereotype threat was activated.

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.¹⁷ 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.

5.2 Within-Sex Studies of Cultural Effects

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.

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.)
Meier-Pesti and Penz, 2008	NSS	—	within females: masculinity-primed vs. femininity-primed	Hypothetical investments and general risk question	30
Ronay and Kim, 2006	NSS	—	within females: with same-sex discussion vs. without	Variety of questions and risk scenarios	50
Meier-Pesti and Penz, 2008	NSS to 0.91	—	within males: masculinity-primed vs. femininity-primed	Hypothetical investments and general risk question	30
Kahan, Braman et al. 2007	.20 to .60	—	white males vs. everyone else	Perceptions of gun, abortion, and environmental risks	2000
Booth and Nolen, 2012	.31 to .71	.68 to .71	within females: single sex vs. coed	Hypothetical investment scenarios, lottery	150
Weaver, Vandello, et al., 2012	.57 to .74	—	within males: gender threat vs. gender affirmation	Gambling experiment	50
Ronay and Kim,	.58 to 1.16	—	within males: with	Variety of	50

2006			same-sex discussion vs. without	questions and risk scenarios	
Carr and Steele, 2010	.68 to 1.05	—	within female: no stereotype threat vs. threat	Experimental gambles	30

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.

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 no detectable effect, on average, 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), as reported in Table 1, 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. They also, however, had some subjects participate in a small group discussion with same-sex peers, while others answered the questions only individually. Though no effect of all-female group discussion was detected within females, 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. 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), to suggest an explanation for these results.

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 (as discussed in Finucane 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). More specifically, it seems that white males who have hierarchical and individualist world views (Kahan et al. 2007) or white males who are also well educated, high income, politically conservative, and who trust authorities (Flynn et al. 1994) are those with, on average, the lowest perceptions of risk in many areas. Flynn, Slovic et al. (1994) 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).

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 $d=1.0$.

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.¹⁸ Others were asked to test a power drill before doing the 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).

In most of the experimental studies summarized in Table 1, no apparent attention was paid to cultural or framing/priming effects, though the results in Tables 2 and 3 suggest that these could be major contributors both to the findings of apparent difference, and to the puzzlingly wide range of substantive differences found. Determining the extent to which findings such as those shown in Table 1

reflect cross-cultural *beliefs* about gender, rather than sex per se, would require new experimental research and/or replication with careful attention to these factors.

6 Conclusion

The statement "women are more risk averse than men" tends to be understood as saying that men and women differ in some substantively important and essential way, by virtue of their sex. A review of the empirical literature, with attention paid to the proper interpretation of statistical results, the quantitative magnitudes of detectable differences and similarities, and the impact of cultural context, reveals that such an understanding is not supported by the actual empirical evidence. Men and women tend to be much more similar in their responses to risk than the popular Mars-versus-Venus understanding would imply, and the role of culture and framing effects may be substantial.

Understanding this point is important for policy purposes, since the perception that there are large, essence-based sex differences in risk-taking and risk-perception have become part of many public and academic discussions. These include discussions about financial market stability (e.g., Kristof 2009); labor market, business, and investment success (e.g., Booth and Nolen 2012; Eckel and Grossman 2008); and environmental policy (e.g., Kahan et al. 2007).

This study also has methodological implications. In regards to future work, the present essay suggests that more attention to issues of context and framing, more attention to the quantitative sizes of differences and similarities, and a more careful interpretation of aggregate results could improve economic research and its

contributions to public discourse. Two specific mathematical tools are supplied: *Cohen's d* and the *Index of Similarity*. These improvements could also be usefully extended to the analysis of differences and similarities in other behaviors, such as competitiveness or criminality, and/or to analysis by categories defined by factors such as age, ethnicity, or nationality as well as sex.

Acknowledgements

Funding for this project was provided by the Institute for New Economic Thinking. Matthew P. H. Taylor supplied expert research assistance. Many authors (as listed in footnotes) graciously shared with us the source data or specific statistics from their studies, for further analysis.

References

- Arano, K., Parker, C., & Terry, R. (2010). Gender-Based Risk Aversion and Retirement Asset Allocation. *Economic Inquiry*, 48(1), 147-155.
- Bakan, D. (1966). The Test of Significance in Psychological Research. *Psychological Bulletin*, 66(6), 423-437.
- Barber, B. M., & Odean, T. (2001). Boys will be Boys: Gender, Overconfidence, and Common Stock Investment. *The Quarterly Journal of Economics*, 116(1), 261-292.
- Baron-Cohen, S. (2003). *The Essential Difference: The Truth about the Male and Female Brain*. NY: Basic Books.
- Barsky, R. B., Juster, F. T., Kimball, M. S., & Shapiro, M. D. (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, D., & Menkhoff, L. (2008). Will Women Be Women? Analyzing the Gender Difference among Financial Experts. *Kyklos*, 61(3), 364-384.
- Bernasek, A., & Shwiff, S. (2001). Gender, Risk, and Retirement. *Journal of Economic Issues*, 35(2), 345-356.
- Blau, F. D., Ferber, M. A., & Winkler, A. E. (2010). *The Economics of Women, Men and Work* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Booth, A. L., & Nolen, P. (2012). Gender Differences in Risk Behaviour: Does Nurture Matter? *The Economic Journal*, 122(558).

- Borghans, L., Golsteyn, B. H. H., Heckman, J. J., & Meijers, H. (2009). Gender Differences in Risk Aversion and Ambiguity Aversion. *Journal of the European Economic Association*, 7(2-3), 649-658.
- Bruhin, A., Fehr-Duda, H., & Epper, T. (2010). Risk and Rationality: Uncovering Heterogeneity in Probability Distortion. *Econometrica*, 78(4), 1375-1412.
- Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender Differences in Risk Taking: A Meta-Analysis. *Psychological Bulletin*, 125(3), 367-383.
- Carr, P. B., & Steele, C. M. (2010). Stereotype Threat Affects Financial Decision Making. *Psychological Science*, 21, 1411-1416.
- Coe, R. (2002). *It's the Effect Size, Stupid: What effect size is and why it is important*. Paper presented at the Annual Conference of the British Educational Research Association, University of Exeter, England,
- Croson, R., & Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, 47(2), 448-474.
- Croson, R., Gneezy, U., & Rey-Biel, P. (Eds.). (2012). *Special Issue: Gender Differences in Risk Aversion and Competition* (Vol. 83(1), Journal of Economic Behavior and Organization).
- Cross, C. P., Copping, L. T., & Campbell, A. (2011). Sex Differences in Impulsivity: A Meta-Analysis. *Psychological Bulletin*, 137(1), 97-130.
- Dindia, K. (2006). Men are from North Dakota, Women are from South Dakota. In K. Dindia, & D. J. Canary (Eds.), *Sex Differences and Similarities in Communication* (pp. 3-18). New York: Routledge.
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences. *Journal of the European Economic Association*, 9(3), 522-550.
- Duncan, O. D., & Duncan, B. (1955). A Methodological Analysis of Segregation Indexes. *American Sociological Review*, 20(2), 210-217.
- Durlak, J. A. (2009). How to Select, Calculate, and Interpret Effect Sizes. *Journal of Pediatric Psychology*, 34(9), 917-928.
- Eckel, C. C., & Grossman, P. J. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization*, 68(1), 1-17.
- Eliot, L. (2009). *Pink Brain, Blue Brain: How Small Differences Grow into Troublesome Gaps--And What We Can Do About It*. Boston: Houghton Mifflin Harcourt.
- Eriksson, K., & Simpson, B. (2010). Emotional reactions to losing explain gender differences in entering a risky lottery. *Judgment and Decision Making*, 5(3), 159-163.
- Factor, R. J., & Rothblum, E. (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, doi:doi: 10.5172/hesr.451.17.3.235.
- Fehr-Duda, H., Gennaro, M. D., & Schubert, R. (2006). Gender, Financial Risk, and Probability Weights. *Theory and Decision*, 60, 283-313.

- Finucane, M. L., Slovic, P., Mertz, C. K., Flynn, J., & Satterfield, T. A. (2000). Gender, race, and perceived risk: The 'white male' effect. *Health, Risk & Society*, 2(2), 159-172.
- Flynn, J., Slovic, P., & Mertz, C. K. (1994). Gender, race, and perception of environmental health risks. *Risk Analysis*, 14(6), 1101-1108.
- Gelman, S. A. (2005). Essentialism in Everyday Thought. *Psychological Science Agenda*(May), 1-6.
- Gneezy, U., Leonard, K. L., & List, J. A. (2009). Gender Differences in Competition: Evidence from a matrilineal and patriarchal society. *Econometrica*, 77(5), 1637-1664.
- Gray, J. (1993). *Men are from Mars, Women are from Venus* NY: HarperCollins.
- Harris, C. R., Jenkins, M., & Glaser, D. (2006). Gender Differences in Risk Assessment: Why do Women Take Fewer Risks than Men? *Judgment and Decision Making*, 1(1), 48-63.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2/3), 1-23.
- Holt, C. A., & Laury, S. K. (2002). Risk Aversion and Incentive Effects. *The American Economic Review*, 92(5), 1644-1655.
- Hyde, J. S. (2005). The Gender Similarities Hypothesis. *American Psychologist*, 60(6), 581-592.
- Joop Hartog, A. F.-i.-C. a. N. J. (2002). Linking Measured Risk Aversion to Individual Characteristics. *Kyklos*, 55(1), 3-26.
- Kahan, D. M., Braman, D., Gastil, J., Slovic, P., & Mertz, C. K. (2007). Culture and Identity-Protective Cognition: Explaining the White-Male Effect in Risk Perception. *Journal of Empirical Legal Studies*, 4(3), 465-505.
- Mistresses of the Universe. (2009, February 8, 2009). *New York Times*.
- Levin, I. P., Snyder, M. A., & Chapman, D. P. (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, G. S., & Säve-Söderbergh, J. (2011). "Girls will be Girls", especially among Boys: Risk-taking in the "Daily Double" on Jeopardy. *Economics Letters*, 112(2), 158-160.
- Martell, R. F., Lane, D. M., & Emrich, C. (1996). Male-Female Differences: A Computer Simulation. *American Psychologist*, 51(2), 157-158.
- Meier-Pesti, K., & Penz, E. (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.
- Nelson, J. A. (2013). The Impact of Gender Stereotyping and Confirmation Bias on Economic Research: Theory and a Case Study. *Working paper, based on "Are Women Really More Risk-Averse than Men?" INET Research Note #12, 2012*.
- Olen, H. (2012). *Pound Foolish: Exposing the Dark Side of the Personal Finance Industry*. NY: Portfolio/Penguin.
- Olofsson, A., & Rashid, S. (2011). The White (Male) Effect and Risk Perception: Can Equality Make a Difference? *Risk Analysis*, 31(6), 1016-1032.

- Olsen, R. A., & Cox, C. M. (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, M., & Ansic, D. (1997). Gender differences in risk behaviour in financial decision-making: An experimental analysis. *Journal of Economic Psychology*, 18, 605-628.
- Reskin, B. (1993). Sex Segregation in the Workplace. *Annual Review of Sociology*, 19, 241-270.
- Rivers, L., Arvai, J., & Slovic, P. (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, R., & Kim, D.-Y. (2006). Gender differences in explicit and implicit risk attitudes: A socially facilitated phenomenon. *British Journal of Social Psychology*, 45(2), 397-419.
- Schubert, R., Brown, M., Gysler, M., & Brachinger, H. W. (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.
- Stanley, T. D., & Doucouliagos, H. (2010). Picture This: A Simple Graph that Reveals Much Ado about Research. *Journal of Economic Surveys*, 24(1), 170-191.
- Stanley, T. D., Doucouliagos, H., Giles, M., Heckemeyer, J. H., Johnston, R. J., Laroche, P., et al. (2013). Meta-Analysis of Economics Research Reporting Guidelines. *Journal of Economic Surveys*, 27(2), 390-394.
- Sunden, A. E., & Surette, B. J. (1998). Gender Differences in the Allocation of Assets in Retirement Savings Plans. *The American Economic Review*, 88(2), 207-211.
- Tanaka, T., Camerer, C. F., & Nguyen, Q. (2010). Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam. *American Economic Review*, 100(1), 557-571.
- U.S. Census Bureau (2010). Current Population Survey - June 2008, Detailed Tables. In P. Division (Ed.).
- Weaver, J. R., Vandello, J. A., & Bosson, J. K. (2012). Intrepid, Imprudent, or Impetuous?: The Effects of Gender Threats on Men's Financial Decisions. *Psychology of Men & Masculinity, Advance online publication.*, 1-12, doi:doi:10.1037/a0027087.
- Wilkinson, L., & Task Force on Statistical Inference (1999). Statistical Methods in Psychology Journals: Guidelines and Explanations. *American Psychologist*, 54(8), 594-604 (591-524 in downloadable document).
- Zakzanis, K. 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.

Notes

¹ Additional 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 cited in this article.

² An extension of Nelson (2013) to include the funnel graph suggested by Stanley and Doucouliagos (2010) is currently being planned.

³ Nelson (2013), however, gives examples of explicit essentialism within the scholarly literature.

⁴ In the literature reviewed here, Eckel and Grossmann (2008) and Dhomen, Falk, et al (2011) are notable for providing any extended discussion of substantive economic significance.

⁵ 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) suggests guidelines that misinterpret the meaning of confidence intervals.

⁶ This definition is, however, complicated by intersex individuals and by those who identify as transsexual or genderqueer (Factor and Rothblum 2008).

⁷ 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.

⁸ 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.

⁹ See previous footnote.

¹⁰ 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, Vandelto et al. (2012).

¹¹ 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).

¹² 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).

¹³ 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).

¹⁴ The articles in Croson, Gneezy et al (2012), in particular, were not yet available at the time this research was being done.

¹⁵ 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 occurred in some of the relatively 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.

¹⁶ 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 et al. 1999).

¹⁷ 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.

¹⁸ 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 losing manhood status in other people's eyes leads them to compensate (or perhaps overcompensate) by taking greater risks and seeking immediate rewards" (Weaver et al. 2012).