As unwitting children of the Enlightenment, we seem to have a Tower of Babel—like craving for absolute certainty. And so both sides in the debate recruit biologists and social scientists as latter-day natural theologians who are supposed to help close the theological gaps by telling us, from a “scientific” perspective, what gender complementarity “really is.” Thus, Recovering Biblical Manhood and Womanhood (RBWM) has chapters on biology, psychology, and sociology, and Discovering Biblical Equality (DBE) has chapters written or cowritten by therapists, a sociologist, and an academic psychologist. But as an academic psychologist and gender studies scholar who did not contribute to either volume, I am now going to try to explain (not for the first time) why this is a misguided exercise. My basic points are these:

(1) Research in neither the biological nor the social sciences can resolve the nature-nurture debate regarding gendered psychological traits and behaviors in humans, let alone pronounce on what gender complementarity “really is.” Thus, Recovering Biblical Manhood and Womanhood (RBWM) has chapters on biology, psychology, and sociology, and Discovering Biblical Equality (DBE) has chapters written or cowritten by therapists, a sociologist, and an academic psychologist. But as an academic psychologist and gender studies scholar who did not contribute to either volume, I am now going to try to explain (not for the first time) why this is a misguided exercise. My basic points are these:

(2) There are very few consistent sex differences in psychological traits and behaviors. When these are found, they are always average—not absolute—differences, and, for the vast majority of them, the small, average—and often decreasing—difference between the sexes is greatly exceeded by the amount of variability on that trait within members of each sex. Most of the “bell curves” for women and men (graphing the distribution of a given psychological trait or behavior) overlap almost completely. So it is naive at best, and deceptive at worst, to make essentialist (or even generalist) pronouncements about the psychology of either sex when there is much more variability within than between the sexes on most of the trait and behavior measures for which we have abundant data.

(3) To adapt one of Freud’s famous dictums, we cannot assume that anatomy is destiny until we have controlled for opportunity. Thus, even when appeals are made to large cross-cultural studies that have found “consistent” behavioral and/or attitudinal sex differences, we cannot assume universality for those conclusions until we have controlled for the existence of differing opportunities by gender across the various cultures.

Let me now address these three points in more detail, after which I will make some modest proposals about how the social sciences might more reasonably be expected to be helpful to both sides in the egalitarian-hierarchist debate.

The limitations of science

Research in neither the biological nor the social sciences can resolve the nature-nurture controversy regarding gendered psychological traits and behavior in humans. The crucial terms here are the words “human” and “psychological traits and behaviors.” First of all, we should not be surprised that, given our creational overlap with all other living organisms (strikingly shown in the various genome projects that are underway) much can be learned about the structure, function, and healing of the human body from animal research models. But without doubt, the most salient biological feature of human beings is the plasticity of their brains. The legacy of a large cerebral cortex puts us on a much looser behavioral leash than other animals, with the result that, more than any other species, we are created for continuous learning—for passing on what we have produced culturally, not just what we have been programmed to do genetically. We are, as it were, hard-wired for behavioral flexibility. Indeed, how could we carry out the cultural mandate to “subdue the earth” (Gen 1:28) as God’s accountable regents if this were not so? And, at the other end of the biblical drama, how could we bring the honor and glory of nations—however suitably cleansed—before God (Rev 21:26) if all the people of all the nations had no more freedom within their common biological form than that which exists in even our closest primate neighbors? And, in between, what would be the point of reading and taking to heart Jesus’ parable of the talents (Matt 25:14–20)?

Ah, yes, some will say, but the biological and social sciences have shown us that men and women have clearly different talents and that these are rooted in biology. Really? Well, let us ask what we have to be able to do in order to conclude that biological sex clearly causes even a small, average behavioral or psychological difference between human males and females. First, we would have to be able to manipulate sex as an independent, experimental variable—that is, randomly assign people to be born with an XX or an XY chromosome apart from all the other specific, genetic baggage they come with. Clearly, we cannot do this: babies come to us as genetic “package deals” who, we should remember, have also had non-random environments for nine months prior to birth. Well, then, perhaps we could take advantage of that marvelous natural experiment known as identical twins, each pair of whom have the same genes, have shared the same uterus, and have been shown to stay pretty similar on many behavioral and psychological measures even when raised in different environments. Surely that says something about the power of biology? Yes, it does—although not as much as one might think— but it explains nothing about the origins of gender differences, because identical twins are always of the same sex.

Perhaps we could randomly assign members of a mixed-sex group of infants to be raised as boys or as girls after they are born, and see just how much they remain stubbornly “masculine” or
“feminine” despite being raised as members of the other sex. But, aside from the fact that this comes close to the sort of science that was done in Nazi Germany, but repudiated in our own society, it would not begin to approximate a double-blind experiment of the sort we use, for example, to test the effectiveness of new medicines, because the cat would be out of the bag (so to speak) as soon as the babies’ caretakers began changing their diapers.7 And, even if we could unambiguously ascertain that boys (for example) are hard-wired to be aggressive or girls are hard-wired to gossip a lot, this would tell us nothing about the desirability of either state of affairs. In a fallen world, we cannot automatically assume that what seems “natural” is thereby desirable by the standards of God’s kingdom. This is a point repeatedly and cogently made by psychologist Cynthia Neal Kimble in chapter 27 of DBE.

So, it is impossible to disentangle biological sex from the other genetic and environmental forces in which it always remains embedded and with which it constantly interacts. This means that the two essential conditions for inferring cause and effect—the manipulation of one factor (sex) and the control of other (biological and environmental) factors—cannot be met. Consequently, “all data on sex differences, no matter what research method is used, are correlational data,”8 and as every introductory social science student learns, you cannot draw conclusions about causality from merely correlational data: “[I]n that sense, it is more accurate to speak of ‘sex-related’ differences than of sex [-caused] differences.”9 So, let us be very clear: when we read about a study—experimental or correlational—that describes obtained, average sex difference of such-and-such a magnitude, that is all it is: a description of the results of a study done in one particular place and time with a particular sample of persons, but unable (even experimentally) to disentangle nature from nurture. It is a description, not an explanation, about the origins of any obtained sex differences.10

Overlap of distributions

On almost all behavioral and psychological measures that have been studied, the distributions (“bell curves”) for women and men overlap almost completely. Ah, yes, some will say, but look how large and consistent those sex differences are—in aggression, nurturance, verbal skills, spatial abilities, and so on. Surely this strongly suggests (even if it cannot absolutely prove) that women and men have innately different talents—“beneficial differences” in the language of both CMBW and (some) CBE adherents. Everybody knows that men are from Mars and women are from Venus—at least on average. Really? Just how large and consistent are such differences after a century of measuring them in domains such as aggression, nurturance, verbal skills, and so on? In other words, just how much do (or do not) those “bell curves” overlap for women and men? Because there is so much bad science journalism floating around about these matters (written by people of every political and religious stripe), some more comments on social science methodology are in order.

I begin with what is known among social scientists as the “file drawer effect.” Since psychology journals began publishing more than a century ago, there has been a heavy bias against accepting studies on males and females that find no statistically significant sex differences. In this kind of research, it appears that no news is bad news for your career, because studies finding no effect for sex are likely to remain unpublished (thus ending up in the author’s file drawer). You can see what this means: even when we do a literature review of many sex-comparative studies (concerning any of the usual suspects: verbal or spatial skills, aggression, empathy, activity levels, etc.) done over many years, our conclusions, at least by the reigning statistical criteria, will be selectively tilted toward finding more rather than fewer sex differences because of the publishing bias I have just described.11

My second and more important point has to do with the misunderstanding that continues to surround the term statistically significant. Another basic methodological caveat is this: a research result that is statistically significant is not necessarily of practical significance. According to the most common tests of significance, if an obtained, average difference between two groups (e.g., women and men doing a math test, volunteer subjects taking an experimental drug versus those taking a placebo, etc.) could have occurred fewer than five times out of a hundred “by chance,” then it is deemed a significant difference. However, with large enough samples and a small enough variability among scores, even a tiny average difference between two groups—i.e., groups whose bell-curve scores overlap almost completely—may be significant in this statistical sense, whereas (because of the file-drawer effect) a much larger average difference that “just misses” being statistically significant will not likely see publication, even though its potentially practical significance may be much greater.12

As a result of such criticisms, a statistical technique called meta-analysis was developed in the 1970s for use in all areas of psychological science, including research on gender.13 As its name implies, this refers to a “super-analysis” that can combine the results of many (e.g., several dozen, sometimes more than a hundred) studies on sex differences in a given domain: aggression, verbal ability, or whatever. This technique differs from earlier ways of reviewing the literature, which simply gave equal weight to all studies examined, tallied how many did or did not show statistically significant sex differences, and came to an “eyeball” or intuitive judgment as to whether reliable sex differences existed in a given domain.14 Instead, meta-analysis converts the findings of a large sample of studies into a common metric known as the average effect size across those studies. This is done not just by “averaging all the average sex differences” across the studies, but also by taking into account the size of each sample and the variability of the scores found in each.15 Meta-analysis allows us to ask, across many studies of sex differences of a certain trait or behavior, just how large that difference (known as d) is, or how far apart the tops of the two bell curves are—the tops representing the place where the male and female mean scores are.16 In other words, across many such studies, just how much do the male and female bell curves (or distributions of scores) overlap?17

As you can see from Figure 1, even when an average effect size (or d) is 1.00 (as was found, for example, in a meta-analysis of
studies comparing self-reported empathy in men and women), the range of scores within each sex is much greater than the average difference between the sexes. But in the many meta-analyses of gender differences that have been done since the 1970s, an effect size (d) even as large as .100 is almost unheard of. Most are in the range from 0.0 (no detectable difference) to .15 (a small difference)—and even the latter means that less than 5 percent of the variability of all the scores can be accounted for by the sex of the participants. This underlines my previous assertion: it is naive at best, and deceptive at worst, to make essentialist pronouncements about either sex when the range of scores within each sex is, for almost all traits and behaviors measured, much greater than the difference between the sexes. (See Figure 2 for some representative meta-analytic results of studies of behavioral and psychological sex differences.)

It gets worse. Meta-analysis is full of embarrassments for gender essentialists, but also for “gender influentialists” who think that even small average sex differences are pregnant with interpersonal, ecclesiastical, and policy implications. For example, as previously noted, the meta-analytic d for women’s versus men’s “empathy” scores based on self-report measures is around 1.00, in the direction of women being more empathetic than men. But, when based on unobtrusive measures (i.e., studies where people do not know they are being measured for empathy), the meta-analytic d shrinks to about .05. One does not have to be a professional social scientist to know what that contrast suggests. Meta-analyses can also be divided according to the particular era in which the studies were done. For example, a meta-analysis of studies of gender differences in verbal fluency done prior to 1973 (when gender roles were more rigidly dichotomized) found an overall, small effect size (d) of .23, in the direction of women scoring higher than men. A similar meta-analysis of studies done after 1973 found an effect size of .11, less than half the size of the earlier one. One also does not have to be a professional social scientist to know that sudden genetic mutations in men and/or women since 1973 are unlikely to have caused such a shift. Genes in humans just do not mutate and spread that fast.

**Attempts to evade these findings**

What do convinced gender essentialists (along with careless science journalists and trendy Mars-Venus advice book writers) do with such findings? The most common strategy is simply to ignore or distort them: to pretend that small, shifting tendencies are absolute gender dichotomies, or something close to it, or to assume that statistical significance is always the same as practical significance. All too many people yearn for simple black-and-white explanations of complex relations, including those involving men and women. (As one of my students memorably observed, “Tendencies don’t sell books.”) A less common strategy nowadays is to pathologize the findings: to claim that, however much those gendered bell curves do—or can—overlap, we have to pull them apart as far as possible in order to approximate God’s (or nature’s, or optimal society’s) “true” purposes for males and females. This was the approach taken by philosopher Jean-Jacques Rousseau in his eighteenth-century educational treatise *Emile*. Rousseau was convinced that “rational, active man” and “emotional, passive woman” were perfect complements for each other. Thus, though he freely conceded that men’s and women’s natural traits were not rigidly dichotomous, he insisted that, if they were not trained to become “opposite sexes,” there was no way they would be attracted to each other and be able to pair bond for life. Two centuries later, this kind of theory was embodied in sociological functionalism, whose adherents maintained that a division of labor by sex—whether or not the corresponding tendencies were enshrined in the genes—was “functional” for the preservation of societies, both past and present, and so should be tampered with only cautiously, if at all.

It is not unheard of for theologians to have taken a similar stance. Abraham Kuyper did so in the early twentieth century, claiming (quite ahistorically and with no clear exegetical warrant) that, however much men’s and women’s capacities “naturally” overlapped, God had ordained, once and for all, that women’s activities be limited almost completely to the domestic sphere.
and men’s to the public arenas of the academy, the church, the marketplace, and the political forum. The woman can lend herself to study [of medicine and law] as well as the man,” Kuyper conceded in 1914. But, he added, because women’s (not men’s) “position of honor” was by divine definition in the home, “whoe
er has man take his place at the cradle and woman at the lect-
ern makes life unnatural.”

So far, the doctrine of separate spheres is not an official affirmation of CBMW’s gender hierarchists, aside from its application to certain church of
ci. But to the extent that gender-hierarchist rhetoric overlaps with romantic Mars-Venus rhetoric, as it does on the shelves of many Christian bookstores, it is a force to be reckoned with in many evangelical churches. And, to the extent that the doctrine of separate spheres, combined with the doctrine of male headship, results in the social and economic disempower-
ment of women (as it has in both preindustrial and industrialized cultures), it does not comport well with biblical notions of jus-
tice. This points to a third strategy, one more frequently invoked in the recent past. Some gender essentialists have reluctantly recog-
nized that neither the Bible nor the natural or social sciences can come definitively to their rescue. Consequently, they take
 refuge in biblically and empirically questionable Jungian gender archetypes and their precursors in Greek mythology and Eastern religions. For example, the 1982 book Let Me Be a Woman warned female Christians readers that Eve, in taking the initiative
to eat the fruit, was trying to be like the “ultimately masculine” God—as if God were somehow metaphysically gendered. The au-
 thor also appealed to the ancient Chinese concept of yin and yang to buttress her “Christian” argument for gender essentialism and gender hierarchy. The author’s brother in a 1978 article frankly acknowledged that the Bible does not supply enough resources to justify talking about God or humans in terms of metaphysi-
cal, eternal gender archetypes. Undeterred by this, he invited his readers to consider the abundance of sexual imagery in pagan myths and came to the conclusion that “a Christian would tend to attach some weight to this.” Really? Why?

Joan Burgess Winfrey is thus right to express concern that “the church may once again opt for Venus-Mars gender rubbish in the interest of cementing roles and putting up divider walls.”

Even if Mars-Venus rhetoric is used only to cement different gen-
der styles rather than roles, it receives virtually no support from the meta-analytic literature which, as we have seen, shows almost complete overlap in the gendered distribution of traits, such as nurturance, empathy, verbal skills, spatial skills, and aggressiv-
ness. The romanticizing and/or rank-ordering of gender archetypes is biblically questionable whether it is done by gender-role traditionalists, by cultural feminists who reverse the hierarchy by valorizing the stereotypically feminine, or by evangelical writers who baptize the trendy Mars-Venus rhetoric with a thin, Chris-
tian-sounding veneer. More in keeping with both the biblical creation accounts of humankind and the overall findings of the social sciences is the bumper sticker that reads, “Men are from Earth, Women are from Earth: Get used to it!”

Figure 2: Some effect sizes (“average ds”) from various meta-analyses of studies of sex differences

<table>
<thead>
<tr>
<th>Trait or behavioral measure</th>
<th>Effect size (average d)</th>
<th>Which sex scores higher?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height (US and UK)</td>
<td>2.00</td>
<td>M</td>
</tr>
<tr>
<td>Throwing velocity</td>
<td>2.18</td>
<td>M</td>
</tr>
<tr>
<td>Empathy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Self-reported</td>
<td>1.00</td>
<td>F</td>
</tr>
<tr>
<td>b) Behaviorally observed</td>
<td>0.05</td>
<td>F</td>
</tr>
<tr>
<td>Helping behavior</td>
<td>0.34</td>
<td>M</td>
</tr>
<tr>
<td>Spatial rotation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Children</td>
<td>0.38</td>
<td>M</td>
</tr>
<tr>
<td>b) Adults</td>
<td>0.54</td>
<td>M</td>
</tr>
<tr>
<td>Interpersonal distance</td>
<td>0.54</td>
<td>F</td>
</tr>
<tr>
<td>Desires many sex partners</td>
<td>0.87</td>
<td>M</td>
</tr>
<tr>
<td>Math computation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Children</td>
<td>0.22</td>
<td>F</td>
</tr>
<tr>
<td>b) Adults</td>
<td>0.00</td>
<td>Neither</td>
</tr>
<tr>
<td>Math problem solving</td>
<td>0.32</td>
<td>M</td>
</tr>
<tr>
<td>Reading skills</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) 5–6 years of age</td>
<td>0.31</td>
<td>F</td>
</tr>
<tr>
<td>b) 11–19 years of age</td>
<td>0.02</td>
<td>F</td>
</tr>
<tr>
<td>Verbal analogies</td>
<td>0.16</td>
<td>M</td>
</tr>
<tr>
<td>Verbal fluency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Studies prior to 1973</td>
<td>0.23</td>
<td>F</td>
</tr>
<tr>
<td>b) Studies after 1973</td>
<td>0.11</td>
<td>F</td>
</tr>
<tr>
<td>Aggression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Children</td>
<td>0.58</td>
<td>M</td>
</tr>
<tr>
<td>b) Adults</td>
<td>0.27</td>
<td>M</td>
</tr>
<tr>
<td>Rough-and-tumble play</td>
<td>0.45</td>
<td>M</td>
</tr>
<tr>
<td>Moral reasoning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) “Care” orientation</td>
<td>0.28</td>
<td>F</td>
</tr>
<tr>
<td>b) “Justice” orientation</td>
<td>0.19</td>
<td>M</td>
</tr>
</tbody>
</table>

Effect size criteria: .20 or less = negligible; .21–.35 = small; .36–.65 = medium; .66 and above = large
olds does not help gender essentialists much, because, even in the animal research on which it is based, the thresholds themselves are variable within male and female subject groups, and the resulting distributions overlap, just as they do for actual behaviors. Moreover, as I noted previously, it is always risky to generalize from animal to human behavior because human brains are structured for much more behavioral flexibility than those of even their closest primate neighbors.

We cannot assume that anatomy is destiny until we have controlled for opportunity

In a final attempt to rescue gender essentialism, some scholars claim that, if a certain gender difference holds up cross-culturally—that is, across many different learning environments—we can more safely conclude that it is “natural” and “fixed.” But this conclusion is also too simple. For example, in chapter 27 of DBE, the author cites (and seems to accept as accurate) cross-cultural studies showing that men “are more oriented toward promiscuity and finding a younger and attractive female partner,” while women are “more concerned with finding older men who have attained financial resources and social status.”

Although she does not reference any of the relevant research, the most-quoted study of this sort is a thirty-seven-nation survey of mate-selection standards by Texas psychologist David Buss. Buss suggested his findings meant that men everywhere are genetically predisposed for reproductive reasons to look for youth and beauty in a prospective mate, while women are more predisposed to look for ambition and wealth in the men they seek to marry. But his study made no attempt to control for the differing opportunities that face women and men in many cultures. That powerful, older men marry gorgeous younger women more than the opposite scenario is certainly the case. But, as New York Times science journalist Natalie Angier wryly observed, “If some women continue to worry that they need a man’s money because the playing field remains about as level as Mars—or Venus if you prefer—then we can’t conclude anything about innate preferences.”

More recently, social psychologists Alice Eagly and Wendy Wood did control for changing opportunities by sex. They took the thirty-seven countries of Buss’s study and rank-ordered them according to two indices of gender equality devised by the United Nations Development Program. One is the Gender-Related Development Index (GDI), which rates each nation on the degree to which its female citizens do not equal their male counterparts in lifespan, education, and basic income (which is still the case, though to varying degrees, in all nations). The other is the Gender Empowerment Measure (GEM), which rates nations on the degree to which women, in comparison to men, have entered the public arena as local and national politicians, and as technicians, professionals, and managers.

Using these two measures, they found that, as gender equality in Buss’s thirty-seven-nation list increased, the tendency for either sex to choose mates according to Buss’s so-called evolutionary sex-selection criteria decreased. Eagly and Wood concluded from this that sex differences in mate-selection criteria are less the result of evolved biological strategies than of the historically constructed sexual division of labor, which makes women dependent on men’s material wealth and men dependent on women’s domestic skills. As this wall of separation breaks down—a process nicely traced by the two UN measures—both sexes revert to more generically human (and might we add, biblical?) criteria to judge potential mates, such as kindness, dependability, and a pleasant personality.

Making relationships the unit of analysis: how the social sciences can help

So far, I have tried to show that the odds are not good for using social science research to define the content of gender complementarity—if by that we mean showing how men and women essentially, or even generally, differ for all times and places. Nor should that surprise us. A responsible reading of Scripture indicates that God has built a lot of flexibility into what we call gender, which is why I always prefer to talk about gender relations rather than using the more static term gender roles. As Richard Hess noted in his treatment of Genesis 1 (chapter 3 of DBE), sex is something we share with other, lower creatures, but gender is a part of the cultural mandate.

If we compare Genesis 1:20–22 with Genesis 1:26–28, we see that God first speaks to both animals and humans in exactly the same terms: “Be fruitful and increase in number and fill [the seas, the earth].” What differs is that the primal human pair is given an additional mandate: to subdue the earth. Reformed theologians have taken this to mean that human beings, whether or not they acknowledge the divine source of this mandate, are called to unfold the potential in creation in ways that flexibly express the image of God, yet stay within the limits of God’s creation norms. What Christians have too often done instead, under the influence of pagan and Greek thought and the doctrine of separate spheres, is to assign men to subdue the earth while telling women to be fruitful and multiply.

This seems to me to get it quite backward. While the cultural mandate does not require a blanket endorsement of androgyny (another example of rigid, ahistoric thinking), it does suggest that any construction of gender relations requiring an exaggerated, permanent separation of activities and/or virtues by sex is eventually going to run into trouble (as it has within the last half century), because such exaggeration is creationally distorted and thus potentially unjust toward both sexes. Sexual dimorphism is indeed part of our creational framework, but gender is something to be responsibly structured and renegotiated throughout the successive acts of the biblical drama—not a mystical, rigid, archetypal given.

Thus, we need to think of gender as much in terms of a verb as a noun: “doing gender” is a responsible cultural activity whose mixed outcomes need to be critically examined in the context of the continuing biblical drama in which we are all actors. For people with a low tolerance for ambiguity, this can be very upsetting. Many of us would rather be like the “wicked and lazy servant” in the parable of the talents (Matt 25:14–30), keeping our assets buried in the cold ground of gender stereotypes and a fall-based gender hierarchy, instead of flexibly multiplying them in the service of God and neighbor.
In chapter 26 of DBE, Jack and Judith Balswick, a sociologist and marriage and family therapist, have perceptively developed a relational approach to gender in the service of just and flourishing marriages. In such marriages, “The locus of authority is placed in the relationship, not in one spouse or the other,” and both independence and interdependence are crucial:

Behind the “two are better than one” Scripture is the idea that two independent persons have unique strengths to offer each other and the relationship. Without two separate identities, interdependence is not possible. Some hold to the notion that dependency or fusion is the ideal . . . [but] two overly dependent persons, hanging on to each other for dear life, have no solid ground on which to stand when things get difficult or an unexpected stress hits.40

At the other, hierarchical extreme, they note, “The dilemma of unequal partnership is that husbands carry the burden of having to know everything and always be right, while wives pretend not to know or suppress what they know is right.”41 In contrast to both these distortions, the Balswicks’ four marital relationship principles—covenant, grace, mutual empowerment, and intimacy—focus less on prescribed roles (which are seen to be flexible and negotiable throughout the family life cycle) and more on processes needed for the ongoing flourishing of couples and families. These include that “partners hold equal status; accommodation in the relationship is mutual; attention to the other in the relationship is mutual; and there is mutual well-being of the partners.”42

Does it matter for these processes that the “partners” are male and female, or does this relations-without-roles model lead to “soft androgyny” and thence to the endorsement of non-heterosexual unions? Clearly not for the Balswicks, since they have included a thoughtful section in their chapter on the demonstrated benefits, for both sons and daughters, of coparenting by fathers and mothers. However, even these gendered and generational dynamics are not as simple as was once thought. Freudian and functionalist theorists believed that boys, for example, needed to have lots of interaction with their fathers in order to learn “correct” masculine attitudes, behaviors, and roles. But there is a wealth of research, both in industrialized and pre-industrial cultures, showing that the more fathers are nurturantly involved with their sons, the more secure those sons are in their gender identity, which is simply the sense of being happy and adequate as a male. At the same time, nurturantly fathered sons are less likely to engage in stereotypical “hypermasculine” behavior, such as antisocial aggression, the sexual exploitation of girls, or misogynist attitudes and actions.43

Similar benefits accrue to nurturantly fathered girls, who are more likely to show independent achievement and less likely to engage in premature sexual and reproductive activity. Why is this so? In cultures and subcultures where fathers are absent or uninvolved, boys tend to define themselves in opposition to their mothers and other female caretakers and to engage in misogynist, hypermasculine behaviors as a way to shore up a fragile gender identity.44 And girls who are not sufficiently affirmed as persons by available and nurturing fathers are at risk of becoming developmentally “stuck” in a mindset that sees sexuality and reproductive potential as the only criteria of feminine success.45 The bottom line appears to be this: children of both sexes need to grow up with stable, nurturant, and appropriately authoritative role models of both sexes to help develop a secure gender identity. But strong coparenting also allows growing children to relate to each other primarily as human beings, rather than as reduced, gender-role caricatures. Paradoxical as it may seem, those who are most concerned to display rigidly stereotypical masculinity and femininity are apt to have the least secure gender identities.

Clearly, this does not require that children’s role models always and only be their biological parents.46 But it strongly suggests there are limits to the diversity of family forms we should encourage around a core norm of heterosexual, role-flexible coparenting, as described by the Balswicks in their DBE chapter. As Genesis 1 reminds us, sex is indeed something we share with the lower animals, and, as such, it is irrelevant to the image of God in humans. At the same time, lifelong cooperation between the sexes is part and parcel—indeed, the climax—of the Genesis 2 creation account in a way that is not required of other animals: “For this reason a man will leave his father and mother and be united to his wife, and they will become one flesh” (Gen 2:24). Sociologist David Fraser notes that this verse holds in tension three essential aspects of marriage: public wedlock (“leaving”), sexual union (“one flesh”), and lifelong covenant (“cleaving”). Yet, he significantly notes, “In this passage the couple is complete without children.”47 Thus, heterosexual pair-bonding is not simply a convenient way to have children, although children are indeed part of God’s promised blessing in creation. It is based on the deeper creational truth that women and men are both created in the image of God, derive equal dignity and respect from that image, and are called to be God’s earthly regents—not separately, nor hierarchically, nor in competition with each other, but cooperatively. This does not mean that all men and women must marry; the New Testament is very clear on the value of singleness. But it does suggest that attempts to form single-sex communities (or to impose a rigid doctrine of separate spheres within families and/or churches) as a way of avoiding the challenges of heterosexual cooperation and gender justice are something less than creationally normative and will eventually be shown to be so by their results.

Notes


3. However, it is of interest to note that, in RBWM, the “sociology” chapter is written by a New Testament scholar and the “psychology” chapter does not review the psychology of gender literature as a whole, but only a small slice of the clinical and developmental literature on gender identity formation, with a heavy emphasis on the author’s own research with one clinical sample of convenience. The “biology” chapter does not review the psychology of gender literature as a whole, and the “psychology” chapter is written by a New Testament scholar.

4. Thus, the file-drawer effect can work either way: it can mask large differences that just fail to attain statistical significance, as well as differences that are neither statistically nor practically significant. Most journals in the psychological sciences only publish about 5 percent of the studies that fail to meet traditional levels of statistical significance, the rest ending up in the file drawers of their researchers. For an accessible discussion of these issues, see Christopher Shea, “Psychologists Debate Accuracy of ‘Significance Test,’” Chronicle of Higher Education (Aug. 16, 1996): 12, 17.

5. Cynthia Neal Kimball mentions this technique in passing in DBE, 473.

6. An example of the use of this earlier method would be the Macoby and Jacklin literature review mentioned in n. 3.

7. Pictorially, the variability of scores refers to how “fat” or “skinny” of the bell curves of the scores are for the groups in any study. To take account of such variability is important, because, other things being equal, the skinnier the bell curves, the less likely it is that an average difference between the groups in the study is due to chance.

8. Note that meta-analysts, unlike those using more standard techniques, do not simply ask, “Did the average difference between the groups—however large or small—manage to make the <.05 cutoff for statistical significance?”

9. This is another way of asking whether the differences between the male and female scores are bigger or smaller than the amount of variability within each sex group, or asking how much of the variance in the scores can be explained by the sex of the participants in the study. The best meta-analyses will include as many unpublished studies as possible (to reduce the file-drawer effect) and also have clear methodological standards for which studies are included, e.g., only studies whose measures have demonstrated construct validity, only studies in which participants are randomly assigned to conditions, etc.

10. A d of 1.00 would mean that, after meta-analysis has been done, the average gap between men’s and women’s scores is a full standard deviation in size. By convention, all “bell curves” or distributions of scores are divided across the curve into eight equal standard deviation units. By convention, effect sizes (d) of 0.00–0.35 are considered small; those from 0.36–0.65 are considered medium, and those above 0.65 are considered large. It is worth noting that, according to one review, 60 percent of the effect sizes found in the psychology of gender are in the “small” range, as compared to 36 percent in all other areas of psychology where meta-analyses have been done. See Janet S. Hyde and Marcia C. Linn, The Psychology of Gender: Advances through Meta-analysis (Baltimore, MD: Johns Hopkins University Press, 1986), and Janet S. Hyde and Elizabeth Asby Plant, “Magnitude of Psychological Gender Differences: Another Side to the Story,” American Psychologist 50, no. 3 (March 1995): 159–61.

11. Good reviews of the meta-analytic research on gender can be found in Lips, Sex and Gender, chs. 3–4, and Vicki S. Hegelsohn, The Psychology of Gender (Upper Saddle River NJ: Prentice Hall, 2002), ch. 3.

12. Jean-Jacques Rousseau, Emile, trans. Allan Bloom (New York, NY: Basic Books, 1979). Rousseau’s idea that the sexes should be “opposite” was one of the first modern departures from the longer-standing Aristotelian notion that women and men were in all ways alike, except that women were “lesser” than men in all their human capacities: for rationality, autonomy, artistry, friendship, etc. For Aristotle, women were—in Dorothy Sayers’s memorable phrase—“The Human-Not-Quite-Human.” See Dorothy Sayers, Are Women Human? (Downers Grove, IL: InterVarsity, 1975), 37–47.


28. Elisabeth Elliot, Let Me Be a Woman (Wheaton IL: Tyndale, 1982).

29. Thomas Howard, "A Note from Antiquity on the Question of Women’s Ordination," The Churchman: A Journal of Anglican Theology 92, no. 4 (1978): 323. Howard is in part following C. S. Lewis’s notion that certain themes in pagan myths (e.g., the dying and rising god) foreshadow the “myth made flesh” in Jesus Christ. But even Lewis realized that such myths are only “a starting point from which one road leads home and a thousand roads lead into the wilderness.” The Pilgrim’s Regress: An Allegorical Apology for Christianity, Reason, and Romanticism (London: J. M. Dent, 1933), 153 (emphasis original). In other words, just because pagan myths were pointing in a Christian direction with regard to their intuitions about dying and rising gods does not make them proto-Christian when they talk about male sky gods and female earth mothers. Lewis himself, however, was clearly inconsistent on this point, embracing as part of “mere” (i.e., basic) Christianity all kinds of assumptions about the “masculinity” of God and essential, metaphysical character differences between women and men. See Candice Fredrick and Sam McBride, Women among the Inklings: Gender. C. S. Lewis, J. R. R. Tolkien, and Charles Williams (Westport, CT: Greenwood, 2001), and Mary Stewart Van Leeuwen, “The AntiReductionist Reductionist: C. S. Lewis, Science, and Gender Relations,” The March 2004 C.S. Lewis Lecture at the University of Tennessee, Chattanooga, http://eastern.edu/academic/trad_undg/sas/depts/psychology/mvanleeu/Lewis.doc.


31. As CBMW appears to do when it says (in Answer 29 to “Fifty Crucial Questions”), “Women are weaker in some ways and men are weaker in some ways; women are smarter in some ways and men are smarter in some ways. . . . God intends for all the “weaknesses” that characteristically belong to man to call forth and highlight woman’s strengths. And God intends for all the “weaknesses” that characteristically belong to woman to call forth and highlight man’s strengths.” CBMW Web site, http://www.cmbw.org/questions/29.php.


38. Even in Buss’s own study, when asked what qualities are most important in a mate, both sexes, on average, ranked love, dependability, emotional stability, and a pleasing personality as the highest four. Only in the average fifth rankings did the differences predicted by Buss’s evolutionary hypothesis emerge. And, as Eagly and Wood showed, those already low-ranking differences were ranked lower and lower as gender equality increased.


42. Balswick and Balswick, “Marriage as a Partnership,” 454.

43. For reviews of this literature, see Scott Coltrane, Family Man: Fatherhood, Housework, and Gender Equity (New York, NY: Oxford, 1996), and Van Leeuwen, My Brother’s Keeper, chs. 6, 8, and 10.

44. This might be grounds for worrying not only about the development of misogyny in boys raised in lesbian households, but boys in conservative Christian homeschooling households, given that almost all such homeschooling is done by mothers. For a sociological analysis of the homeschooling movement in America, see Mitchell Stevens, Kingdom of Children: Culture and Controversy in the Homeschooling Movement (Princeton, NJ: Princeton University Press, 2001).

45. Scott Coltrane’s analysis of almost a hundred preindustrial societies (note 43) shows that nurturant fathering of children also correlates strongly with reduced abuse of women and greater empowerment and voice for women in the cultures where involved fathering takes place.

46. In fact, given that the metaphor of adoption is such a central one in the overall biblical narrative, I am surprised that neither RBMW nor DBE has a chapter on its significance for the organization of family and church life. See, for example, Jeanne Stevenson-Moessner, The Spirit of Adoption: At Home in God’s Family (Louisville, KY: Westminster John Knox, 2003), and Timothy P. Jackson, ed., The Morality of Adoption: Social-Psychological, Theological, and Legal Perspectives (Grand Rapids, MI: Eerdmans, 2005).