CHAPTER 10

GENDER ROLES AND HEALTH

Karina Davidson-Katz

In Western society today, women report more physical illness, more psychological distress, and more psychiatric symptoms than do men. Yet women live longer than men. Most of us assume that these sex differences exist across times and cultures, and have biological causes.

But are these assumptions warranted? The sex difference in mortality is greater or lesser, or even reversed, depending on what culture is examined (Waldron, 1982). The sex difference in illness reports changes size with the method of data collection (Nathanson, 1978). The extent of the sex difference in psychological distress fluctuates with this century’s progression (Kessler & McRae, 1981). Finally, whether there is a sex difference in psychiatric symptomatology depends on environmental factors such as education (Jenkins, 1985). Thus, contrary to our assumptions, it appears that sex differences in morbidity and mortality are not determined exclusively by biology.

Even when a sex difference is consistently found—as in the case of depression—biology alone cannot provide a complete explanation (Nolen-Hoeksema, 1987). To fully explain sex differences in health, one needs to seek more than just biological determinants; one also must ask what psychological variables both differentiate women and men and predict good health outcomes. One such variable that is being investigated is the gender role adopted by the individual (Maccoby & Jacklin, 1974).

What is a gender role? The term has been used to refer to attributes, preferences, characteristics, stereotypes, expectations, and behaviors. This kind of multifaceted definition does not lend itself well to the investigation of gender roles and outcome variables. Thus, more precise and specific definitions of terms are in order.

Widom (1984) suggested dividing the study of sex differences into biological, psychological, and social components. Sex differences examined at the biological level will be referred to here simply by the term sex. Next, sex differences can be investigated at a psychological level; individual differences in characteristics, behaviors, preferences,
and self-concept are then the variables of interest. The psychological category of sex differences will be referred to here as gender roles. Masculinity refers to the psychological characteristics associated with males' gender role; femininity to the psychological characteristics associated with females' gender role. When theorists intend to refer only to one component of a gender role, such as expressive behavior, the element of interest will be separately named. This separation should help the reader avoid invoking all connotations of the term gender role when only one part is under investigation. Finally, sex differences can be examined at a social or cultural level. Shared beliefs about what constitute appropriate behaviors, characteristics, abilities, and preferences for each sex will be called gender role stereotypes.

These three divisions of sex differences are not entirely independent of one another; gender role stereotypes in conjunction with one's sex likely influence the gender role that one eventually adopts. For example, being female may not directly cause passivity and emotional expressivity, but it may do so indirectly by activating a socialization process in which a girl internalizes the gender role stereotype of her culture—passivity and expressivity, in this example.

Distinguishing among the various meanings hidden in prior uses of the term gender roles allows one to examine separately biological, social, and psychological theories. This chapter will review primarily the last set of theories.

Although most of the current psychological and social theories linking gender roles and health have received some empirical support, none of them adequately explains how gender roles cause good (or poor) health. I will begin my discussion by presenting the various gender role theories. Then I will highlight the limitations of these theories by analyzing the difficulties with current research. Next, I will describe two new research directions that may allow us to explore the causal relationship between sex-differentiating personality variables and health. Finally, I will present some of the evidence produced so far by researchers who have pursued these new avenues.

**EVOLUTION OF GENDER ROLE THEORY**

**Traditional Theories**

The view prevalent prior to 1960 was that good health outcomes are linked with sex-appropriate gender role affiliation (Spence & Helmreich, 1978). Early gender role theorists contended that assuming the gender role appropriate to one's biological sex carries psychological, social, and evolutionary benefits. Adoption of a biologically inappropriate gender role was thought to be the result of intrapsychic conflict or improper socialization and to cause distress and pathology in both the individual and society.

One of the major assumptions of these early gender role theorists was that masculinity and femininity are polar opposites; the more you possessed attributes of the other gender role, the less you possessed attributes of the other gender role. Early scales that measured gender role affiliation tended to reflect this assumption (Constantinople, 1973).

Although a few early theorists proposed that masculine and feminine attributes—particularly expressiveness and agency—may not be polar opposites (Bakan, 1966), no serious interest in the relationship between masculinity and femininity occurred until the 1970s.

**Bem’s Theories of Androgyny**

In 1974, Bem offered an alternative hypothesis concerning the relationship between good psychological health and gender roles. She also introduced a new measure of gender role affiliation—the Bem Sex Role Inventory (BSRI). Bem suggested that those individuals scoring either high in both masculine and feminine gender role affiliation or low in both (called androgynous) are better off than those scoring high only in the gender role associated with their biological sex (called sex-typed). She suggested that the ability to perform both masculine and feminine behaviors is healthy, regardless of one's sex. Androgynous individuals, according to Bem, “might be both masculine and feminine, both assertive and yielding, both instrumental and expressive—depending on the situational appropriateness of these various behaviors” (p. 155).

Although this theoretical formulation of androgyny and its relationship to psychological health may seem straightforward, it is not. As critics later suggested, there are a number of hidden assumptions and logical problems with Bem’s original theory.

For example, although Bem was clear in stating that androgynous individuals should have better health than others, she was not definite about how or why this relationship operates. Other the-
orists presumed Bem proposed that since androgy nous individuals have both masculine and feminine behaviors in their repertoire, they can respond to stressors with the most adaptive coping behavior. Hidden within this interpretation is an important assumption; namely, that the androgy nous person can identify the behavior that is most appropriate in any situation. That is, it is not simply the possession of masculine and feminine attributes that leads to good health, but also the ability to use the attributes judiciously. For example, faced with a friend whose spouse has died suddenly, a stereotypical masculine response might be to offer help with the funeral arrangements. A stereotypical feminine response might be to offer sympathy and attention to the bereaved friend. Bem would have to say that the androgy nous individual could not only provide both kinds of assistance but also select the kind that would be most helpful at a given moment.

It should be noted once more that Bem did not explicitly state this assumption about the ability to select the right behavior; she merely implied it when she stated that androgynous individuals use masculine and feminine attributes “depending on the situational appropriateness of these various behaviors” (p. 155). The theory implicit in these comments has been labeled the emerging properties theory of androgyny, because it suggests that when masculine and feminine traits are both present, new qualities, over and above the masculine and feminine qualities, emerge (Hall & Taylor, 1985). These new qualities are flexibility and adaptability.

It is possible that Bem meant only that the simple presence of both types of attributes is better than possessing one set of attributes; this theory is different from the emergent properties theory and has different implications for the relationship between androgyny and health. This variation of the androgyny theory had been labeled the main effects theory (Taylor & Hall, 1982). Most researchers, however, understood Bem to imply the emergent properties theoretical assumptions, and some of Bem’s later articles (e.g., Bem, 1975) more explicitly stated the assumptions necessary for the emergent properties theory.

Bem (1974) also stated that the key to androgyny is the balance of reported masculine and feminine attributes, regardless of their number. In other words, people with no gender role-related attributes and persons with many attributes of both gender roles were grouped together under the rubric androgyny. The implication was that they are equally better off than sex-typed individuals because their choice of behavior is not restricted by gender role stereotypes. This way of conceptualizing androgyny has been labeled the balance theory of androgyny (Taylor & Hall, 1982).

If it appears that there are as many as three different theories of androgyny that may have been forwarded in the Bem (1974) article, with three different implications about who is healthy, this is indeed the case.

Despite the untested (and often unstated) theoretical assumptions inherent in the proposed relationships between gender roles and health, claims that androgyny is an index of psychological health were quickly accepted; therapy that widened a client’s behavioral repertoire to include both masculine and feminine behaviors came to be considered desirable (Gilbert, 1981). Those who scored high only on the scale associated with their biological sex were called “restricted” and “psychologically inferior” to the androgynous types (Kaplan & Bean, 1976).

However, there was also criticism of Bem’s original theory and measurement of gender roles. Spence, Helmreich, and Stapp in 1973 introduced their own measure of gender roles, the Personal Attributes Questionnaire (PAQ). While presenting their own measure, they pointed out that Bem’s measurement of androgyny captured a heterogeneous group of individuals. In contrast, Spence and her colleagues classified their subjects as above or below the median on the two gender role dimensions and thus created four types of gender roles: masculine (high on masculinity and low on femininity), feminine (high on femininity and low on masculinity), androgynous (high on both scales), and undifferentiated (low on both scales). They found that both masculinity and femininity had main effects on a measure of self-esteem, and the effects were additive, so that high masculine, high feminine scorers had the highest self-esteem. Thus, individuals whom Bem would have included in the androgynous category (specifically, those who scored low on both scales) had significantly lower self-esteem than the androgynous individuals who scored high on both scales.

Although some have seen this revision as merely a difference in scoring systems, Taylor and Hall (1982) pointed out that it implies a different theory of androgyny; with this scoring change the balance theory of androgyny is discarded in favor of the main effects theory.

Bem initially saw this new scoring procedure as competing with her own (Bem & Lenney, 1976),
but she eventually revised her criteria for androgy-
ny so that the category included only those who
scored high in both masculine and feminine at-
tributes; those individuals low in both were no
longer considered androgy nous (Bem, 1977). In
accepting this revision to her scoring scheme, in
effect, Bem was changing her theory from the bal-
cance to the main effects theory of androgy

Testing Theories of Androgy

Taylor and Hall (1982), in their excellent disen-
tanglement of the various theories of androgy
ny, pointed out that the balance and main effects the-
ories can be tested against each other and do not
require completely different methodologies as
originally thought. They suggest entering masu-
culinity scores (M), femininity scores (F), and their
interaction (M+F) into a two-way analysis of vari-
ance (ANOVA). The balance theory would be sup-
ported by a significant interaction term in a two-
way ANOVA, whereas the main effects theory
would be supported by two significant main ef-
fects in a two-way ANOVA. According to Hall and
Taylor (1985) all three effects (i.e., each of the M,
F, and M+F terms) must be significant for the
emergent properties theory of androgy nous to be
supported. In this way, different conceptions
of androgy nous can be tested against each other with
the same data.

Hall and Taylor (1985) went on to point out
some theoretical confusion that theorists have
generated in setting up tests of the three androgy
ny theories. Hall and Taylor believe that many
investigators forwarding the main effects theory
have really been discussing the emergent prop-
ties theory. That is, main effects theorists have
been postulating that special benefits accrue to
individuals who score high on both the masculin-
ity and femininity scales, but they have tested this
claim with analyses appropriate for the additive
effects model. Thus, there is a history in gender
role research of explaining or forwarding one the-
y of androgy nous, and testing another. This prac-
tice has added conceptual as well as methodologi-
cal confusion to the entire area.

Even though the three theories of androgy
ny can be tested against one another as described,
most researchers have preferred to follow Bem’s
(1977) lead and test only the main effects theory
of androgy nous. Hence, they do not offer the in-
formation that would be necessary for the other the-
etical formulations to be assessed. For this rea-
son, I will examine primarily the main effects the-
ory; the few studies that do explicitly test the three
theories against one another will be noted.

GENDER ROLES AND
PSYCHOLOGICAL HEALTH

Extensive research on gender roles has pro-
duced little evidence that androgy nous is advan-
tageous to psychological health (Pedhazur &
Tetenbaum, 1979; Taylor & Hall, 1982). In this re-
search, psychological health has been measured
by a number of indexes, including overall adjust-
ment, ego strength, self-esteem, psychiatric symp-
toms, depressive symptoms, and level of anxiety.

Sufficient research has been conducted on the
relationship between psychological health and gen-


Bassoff and Glass (1982) analyzed 26 studies
and found a small, significant relationship be-
tween masculinity and psychological health, the
latter measured with self-esteem and depression
scales. No significant relationship between femi-
ninity and health was found. Thus, the main ef-
fects theory of androgy nous was not supported; sup-
port would only have been shown if both the
masculine and feminine scales were significantly
associated with psychological health.

In a second meta-analysis, Taylor and Hall
(1982) concluded that there was not much support
for either the balance or the main effects androgy
ny theories, whereas there was support for a mas-
culinity effect across a number of psychological
health measures. Hall and Taylor (1985) pointed
out that if there is no support for either the bal-
cane or the main effects theory, there is clearly no
support for the emergent properties theory (which
requires support for both of the others).

Whitley (1983) performed a meta-analysis of 35
studies investigating the relationship between self-
estee m and gender roles. Although he found weak
support for both the main effects and balance the-
ories of androgy nous, he pointed out that androgy
ny, as defined by either theory, accounted for only
1% of the variance in self-esteem; masculinity
alone explained approximately 27% of the vari-
ance.

In a subsequent meta-analysis, Whitley (1984)
explicitly tested the three androgy nous theories (bal-
cane, main effects, and emergent properties)
against one another. He found that none of the
three variations of the androgy nous hypothesis was
supported; once again, only the masculinity scale predicted meaningfully to psychological health.

Meta-analyses are not always considered legitimate summaries of findings because they treat crude, uncontrolled studies and sophisticated, carefully planned studies as equals (Searles, 1985). But even individual studies that are methodologically sound and theoretically rigorous have not offered support for any of the androgyny hypotheses.

For example, Nezu, Nezu, and Peterson (1986) examined the moderating influence of gender roles and family support on the relationship between negative life stress and depressive symptomatology. They found that the more masculine one was, the lower the reported depression under high levels of stress. Femininity (and the interaction between masculinity and femininity) was not significantly related to depression.

Using a longitudinal design, Roos and Cohen (1987) examined the buffering effects of social support and gender roles on the relationship between life stress and trait anxiety and depression. They measured all variables cross-sectionally and then measured all variables but gender role orientation 8 weeks later. They found, with both the cross-sectional and the longitudinal data, that masculinity was a significant buffer of negative life events, whereas femininity was not. To the author's surprise, however, the interaction between masculinity and femininity (in this case, androgyny as defined by the balance theory) was also a significant buffer of stress when mood was measured 8 weeks later.

GENDER ROLES AND PHYSICAL HEALTH

There has been relatively little research into the nature of the relationship between gender roles and physical health. One reason for this is that gender role theorists usually do not explicitly state that gender roles should be related to physical health. A second reason for this dearth of relevant research has been the difficulty in defining optimal physical health. In the studies that will be briefly reviewed here, three indexes of physical health were used: self-perceived health status, use of medical resources, and physiological reactivity to stress.

Olds and Shaver (1980) found that high scorers on the femininity dimension reported significantly higher numbers of physical symptoms than all others. Consistent with this result was Hatzenbuehler and Joe's (1981) finding that masculine males reported fewer health problems than other males, as well as Wech's (1983) finding that masculine sex-type females reported significantly fewer health problems than feminine females. But it is unclear from such findings whether masculine types are healthier or just more reluctant to admit to physical failings (Meinecke, 1981). The only study to find androgynous individuals (as defined by the main effects theory) reporting better health than their masculine peers was a study by Downey (1984), which had middle-aged males as its subjects.

Another index of physical health that has been measured is (infrequent) use of physician services. Marcus and Siegel (1982) found that high scorers on the masculine dimension made less use of medical resources than other gender role types.

Finally, some researchers have used cardiovascular and other physiological responses to lab stressors as health indexes, because greater reactivity has been associated with greater vulnerability to long-term health problems. Some prospective studies have supported this hypothesis, particularly when cardiovascular reactivity and later cardiac problems are examined. There is also some evidence that there is a link between short-term reactivity and certain personality variables (Krantz & Manuck, 1984; Julius, Weder, & Hinderliter, 1986).

Many psychological variables have been examined as possible causal agents of hyperreactivity to lab stressors, but gender roles have rarely been included. Myrsten, Lundberg, Frankenhaeuser, Ryan, Dolpin, and Cullen (1984) measured the gender role of their subjects and then recorded blood pressure and electrocardiograph (ECG) responses to an achievement stress and an orthostatic stress (which involves placing subjects on a horizontal tilt-table and then tilting them feet downward by 45 degrees). They found that there was no significant sex effect on the "stressed" blood pressure, but that there was a significant sex-times-gender role interaction. Post-hoc analyses of their results indicated that among males, feminine types had the lowest blood pressure during the stressors, whereas among the females the undifferentiated types had significantly lower blood pressure than the other females.

Although there have been no other direct examinations of gender roles and reactivity, Frankenhaeuser and her colleagues have done some in-
direct work in this area. Looking at neuroendocrine responses to an achievement stressor, Frankenhaeuser, Lundberg, and Forsman (1980) found that males were far more reactive than females. In subsequent work employing similar measures, Frankenhaeuser (1982) split her females into traditionally employed and nontraditionally employed. Results indicated that the latter group was not significantly different from the males in reactivity to achievement demands; of the females, only the traditionally employed group was significantly less reactive than the males.

Other researchers have examined the effects of gender role–related interests on levels of epinephrine secretion; high levels over a long period are thought to contribute to cardiovascular disease (Polefrone & Manuck, 1987). Girls showing the smallest elevations in urinary epinephrine during an exam were more oriented toward a traditional feminine role than were girls showing higher elevations (Rauste-von Wright, von Wright, & Frankenhaeuser, 1981). During an achievement task, female undergraduates who had masculine academic interests were closer to males in level of urinary epinephrine excretion than were females who did not have masculine academic interests (Collins & Frankenhaeuser, 1978).

Thus, there have been few tests of any androgyny hypothesis in the area of physical health. When self-report measures of health have been used, none of the androgyny hypotheses have been supported; only masculinity appears to predict good health. In the area of short-term reactivity, there is a (very tentative) suggestion that higher scores on the masculinity dimension predict greater reactivity, which in turn is associated with greater vulnerability to long-term health problems. In all but one of the reactivity studies, gender roles were not measured with the usual measures—instead, variables such as masculine academic interests were used. So although these results are thought provoking, conclusions about gender roles and reactivity should await further investigation (Polefrone & Manuck, 1987).

**MASCULINITY THEORY OF HEALTH**

In response to the above findings, a fifth theory has been proposed. According to this theory, called the masculinity theory, the higher one's masculinity gender role score, the better off physically and psychologically one will be, regardless of one's sex or femininity score (Antill & Cun-ningham, 1979; Silvern & Ryan, 1979). Some theorists have proposed that masculinity is even more beneficial for females than it is for males (Jones, Chernovetz, & Hansson, 1978).

These findings also have suggested to some investigators that increasing masculinity would be an appropriate and desirable goal of psychotherapy (Vogel, 1979). Given the psychometric, methodological, and theoretical weaknesses of the research performed to date, however, such a conclusion seems premature at best. These weaknesses will be outlined in detail in the following section.

**LIMITATIONS OF GENDER ROLE RESEARCH**

**Correlational Nature of the Research**

Plaguing the entire investigation of the relationship between gender roles and health is the problem that almost all studies have been correlational in nature. For example, shared methodology could in fact be responsible for the previously noted relationship between masculinity and health. There are two possible ways of minimizing potential artificial findings of correlational designs: randomizing assignment to conditions or using more sophisticated research designs. Gender role assignment is obviously impossible to randomize, but very little effort has been made to improve correlational designs by implementing prospective or longitudinal data collection.

As mentioned earlier, Roos and Cohen (1987) conducted one of the few studies in which longitudinal data were collected on gender roles and psychological health. In support of the masculinity hypothesis, the investigators found that masculinity was the only significant buffer between life stress and mood disturbance when all measures were obtained at the same time. However, when the longitudinal data were examined, the balance theory of androgyny was supported. That is, masculine and feminine subjects became more depressed in response to high life stress than did androgynous and undifferentiated subjects.

Even though these findings suggest that there may be some support for the balance theory when measurement artifact is minimized, all correlational studies, including longitudinal ones, are vulnerable to alternative causal explanations. These findings do, nevertheless, allow us to ques-
tion the support for the recent, empirically derived masculinity theory.

Social Desirability Confound

A second criticism about gender role research has concerned a possible social desirability confound, or the tendency to endorse the socially approved answer rather than the socially disapproved one. With the self-report instruments typically used to assess gender role identification, it is clear that there is the potential for social desirability to play a large role in gender role assignment.

Bem (1974) employed two different methods to ensure that her measure was not simply tapping social desirability. She first had students rate each item for its social desirability for the appropriate sex. Bem reported that the average social desirability ratings of her sex-appropriate traits were not significantly different: her masculine items for males were as socially desirable as her feminine items for females.

Second, in the original gender role scale Bem included a set of "neutral" socially desirable items, to ensure that high masculine, high feminine individuals were not simply endorsing socially desirable items. With these two methods in place, Bem declared that her gender role measure was not greatly influenced by social desirability.

Pedhazur and Tetenbaum (1979) attempted to replicate Bem's findings concerning the lack of influence of social desirability on gender role assignment and discovered two problems. First, some of the feminine items were actually socially undesirable with their sample of subjects—a finding also reported by Silvern and Ryan (1979). These studies suggest that there are sample-to-sample fluctuations in the relative social desirability of the M and F scales.

Next, Pedhazur and Tetenbaum asked their subjects to rate the social desirability of the items for an adult of unspecified gender and found that the feminine traits dropped considerably in social desirability. One might argue that when no gender is announced, subjects assume that the adult is a male; however, the work by Broverman, Broverman, Clarkson, Rosenkrantz, and Vogel (1970) suggested that there are other influences to consider. They found that mental health professionals rated certain traits as mentally healthy for women, others as mentally healthy for men, and that these resulting stereotypes mapped quite well onto their stereotypes of mentally unhealthy and mentally healthy adults, respectively. That is, "healthy" adults (with no sex specified) have masculine traits, and "unhealthy" adults (again, with no sex specified) have feminine traits. This finding suggests that it is unhealthy, or socially undesirable, to be viewed as feminine.

Whitley (1983) also raised the question about shared variance between gender role scales and social desirability. He noted that the M and F scales of both the BSRI and the PAQ ask subjects to rate the degree to which they have socially desirable—but not socially undesirable—traits. He also pointed out that most self-esteem measures ask subjects to do exactly the same thing: to endorse a number of socially desirable items. Thus, it is not surprising that masculine items (which are more socially desirable) correlate significantly with self-esteem items (which are also socially desirable). What is not clear is whether this relationship is due to the causal influence of masculinity on self-esteem or to some third variable, such as social desirability, which is likely to be associated with higher scores on both masculinity and self-esteem.

If social desirability is related to masculinity, then the relative influences of the two overlapping constructs in determining self-esteem must be ascertained. Is it masculinity per se or the part of masculinity that overlaps with social desirability that is responsible for the increase in reported well-being? A revised gender role measure that could isolate the effects of social desirability and gender roles would help disentangle this knot.

In order to redress the possible social desirability confound in the PAQ, Spence, Helmreich, and Holahan (1979) extended their measure (hereafter called EPAQ) to include one scale of socially undesirable masculinity items and two scales of socially undesirable femininity items. By using such a scale, in which endorsement of both socially desirable and socially undesirable items determines gender role assignment, one can discover whether it is simply the endorsement of any masculine items (regardless of their social desirability level) that is related to self-esteem or whether only socially desirable masculine items predict high self-esteem.

It should be noted that Spence and her associates also created a fifth scale when analyzing the factor structure of their new gender role measure. The items on this new scale were bipolar; the ideal man fell on one side of the scale and the ideal
woman fell on the other. The possible meaning of this new scale will be discussed below.

**Use of Masculine-Dependent Measures**

A third criticism about gender role research is that the benefits of masculine traits are more behaviorally anchored and thus more likely to be studied than the benefits of feminine traits, which are more relational (Hall & Halberstadt, 1980). If many of the outcome measures in previous studies were from the masculine domain, then it is not surprising that masculinity was significantly related to them.

Indeed, one could argue that much of the previous research cited concerning the gender roles–health links has instead tested the construct validity of the masculinity scales. Even such constructs as self-esteem and locus of control (for achievement) can be viewed as more masculine than, for example, empathy or emotional expressivity.

Taylor and Hall (1982) point out that in order to test whether or not androgyny is related to health, one should have outcome measures that are not sex typed or from one gender role domain. For example, evidence that androgynous individuals (as defined by the emergent properties theory) are superior to all others on non–sex-typed outcome measures would be considered support for an androgyny hypothesis. However, evidence that androgynous and feminine individuals are superior to others on female-typed outcome measures is simply construct validation of the femininity scale; similarly, evidence that androgynous and masculine individuals are superior to others on male-typed outcome measures is again only construct validation of the masculinity scale.

In one of the studies that examined a clearly nonmasculine dependent measure (marital satisfaction), Antill (1983) found that subjects’ femininity was significantly related to their own increased marital satisfaction. In addition, they found that partners who had a feminine partner (regardless of the sex of that partner) were significantly happier than partners who had other types of partners. Similarly, femininity was positively and meaningfully related to a number of female-typed outcome measures in a meta-analysis of such studies (Taylor & Hall, 1982). Although these are not a substantive test of a gender roles–health relationship—they are again construct validity studies—they do suggest that there may be some benefits accruing to those who are feminine.

Perhaps these benefits are not being studied because the feminine outcome measures are less amenable to measurement.

**Multidimensionality of the Bem Sex Role Inventory**

As mentioned earlier, the most popular instrument for determining gender role is the BSRI. There are two assumptions about Bem’s scale that need to be discussed.

The first assumption is that masculine and feminine gender roles are independent; individuals can have both, neither, or some combination of the two. Bem has been quite clear on this point, stating that her scale was designed to measure the two gender roles as independent dimensions (Bem, 1974). In factor analytic terms, this means that two orthogonal factors are hypothesized.

The underlying factor structure of the BSRI has been found to deviate substantially from the factor structure proposed by Bem. Pedhazur and Tetenbaum (1979) conducted a study in which they asked students to rate the average man and woman with the BSRI items, using the same instructions Bem (1974) had employed. They then gave the scale to a second sample and asked this new sample to rate themselves (and not stereotypes) using the BSRI. Comparison of factor analyses of these two studies revealed clearly different factor structures, neither of which had two factors. Thus, the assumption that the BSRI consists of only two orthogonal factors was not supported.

Wilson and Cook (1984) examined the concurrent validity of four popular gender role instruments, including the BSRI and the EPAQ. When examining the factor structure of the BSRI, they found 10 factors with eigenvalues exceeding 1. Only the EPAQ had the theorized number of factors—5. This finding also supports the contention that the BSRI does not measure two orthogonal dimensions.

A second assumption underlying Bem’s measure is that both the M and F scale are unidimensional. That is, feminine traits such as affectionate and warm should correlate positively with each other as well as with other feminine traits such as innocence and yielding. If this assumption is correct, then the various items belonging to one gender role should load on only one factor.

Again, however, factor analyses reveal that neither the M nor the F scale assesses a unidimensional cluster of attributes. Wilson and Cook
GENDER ROLES AND HEALTH

(1984) reported that the BSRI M scale yields four factors with eigenvalues above 1, and the BSRI F scale yields five factors; only the EPAQ M and F scales yields one factor each. It appears then that the BSRI M and F scales are neither orthogonal nor unidimensional.

It is interesting to note that the BSRI item *masculine* itself does not load on the same factors as the items from the BSRI M scale; nor does the BSRI item *feminine* load on the same factors as the items in the BSRI F scale (Pedhazur & Tetenbaum, 1979). Instead, these two items load on a single (distinct) factor in a bipolar fashion. What, then, are the M and F scales measuring?

**Predictive Validity of the Bem Sex Role Inventory**

Spence and Helmreich (1978) pointed out that many gender role theorists make the questionable assumption that there is a high correlation between gender roles (i.e., self-ascribed traits) and behaviors. In so doing, they raised a serious concern about the predictive validity of the BSRI. Whitley (1988) conducted a multitrait-multimethod analysis of masculinity, femininity, and self-esteem to explore the predictive validity of masculinity and femininity. He collected (self-reported) trait and behavior measures of all three variables and found that masculine trait measures and self-esteem overlapped to such an extent that the two were indistinguishable. In fact, masculine trait measures were more related to self-esteem than they were to measures of masculine behavior. Given this finding, it is not surprising that masculine gender role scores predict to self-esteem; the problem is that masculinity gender role scores may not predict to masculine behavior.

**LIMITATIONS OF GENDER ROLE THEORY**

Given the factorial complexity of masculine and feminine gender role measures and the lack of extensive validity testing, one must ask what is being measured when one obtains a gender role score for an individual.

As previously mentioned, the term *gender role* has many meanings. It can encompass any personality traits, behaviors, preferences, occupational categories, or social positions that are associated with one sex more than the other. Psychologists who theorize about gender roles are typically concerned with personality traits and behaviors, rather than the panoply of variables associated with one sex more than the other.

But what if the constellation of personality traits and behaviors said to be associated with a gender role change as a result of whom you ask and when you ask? An operational definition of a construct that fluctuates according to culture and time should be disquieting to investigators.

What then do theorists mean when they say that someone is masculine or that someone else is feminine? Spence (1985) posed this question and arrived at the unsettling conclusion that, historically, theorists have measured (and therefore appear to have meant) simply those attributes that differentiate the sexes. If sex differences are defined in this way, even height can be considered an indicator of gender role, because men are generally taller than women. Clearly, contemporary gender role theorists thought that they were successfully limiting the range of the term by not using all sex-differentiating qualities as indicators of a gender role. However, this restriction has not led as hoped to a less ambiguous concept. What is still sorely lacking is an explanation of what the terms *masculinity* and *femininity* denote.

If gender role researchers were to limit their conclusions to such statements as "These particular sex-differentiating items are related to this health outcome measure," then the unspoken assumptions and connotations that cleave to the concepts of masculinity and femininity could be avoided. However, gender role theorists typically believe that they are measuring more than simply sex-differentiating traits. Often they propose that they are examining the amount or degree of gender role affiliation the individual possesses, or how sex typed the individual is. Thus, they claim to be measuring more than how much of certain sex-differentiating traits an individual possesses.

Although gender role theorists have not defined masculinity and femininity without ambiguity, lay individuals in our society (who use the terms so frequently) may be able to do so. Myers and Gonda (1982) asked 747 community residents and 233 college students to tell them what the terms *masculine* and *feminine* mean, using an open-ended response format. The authors were able to code the responses into four broad categories: personality or behavioral characteristics, gender (sex) reference, physical reference, and societal or biological reference. Gender and physical characteristics were given more frequently than personality
and behavioral characteristics. In fact, Myers and Gonda noted that only 10% of the community responses and 14% of the college responses reflected the item content of the BSRI. They also noted that there was no consensus among subjects about what the terms specifically mean.

It is commonly assumed in Western society that females are more inclined to have an expressive, communal orientation, whereas males are more apt to have an instrumental, assertive orientation. In fact, the EPAQ, and to some extent the BSRI, appear to measure these two factors, rather than global masculinity and femininity. (The actual items masculinity and femininity, it will be recalled, load on another, bipolar factor.) Psychometric and theoretical analyses thus indicate that current gender role measures actually tap (at best) an expressive, communal orientation and an instrumental, assertive orientation. The EPAQ accomplishes this task more successfully than does the BSRI, which appears to tap many other factors as well.

All the other traits and behaviors that are thought to differentiate the sexes in Western society have not yet been reliably identified and measured, and so cannot (yet) be discussed. Therefore, rather than referring to the two factors assessed by gender role measures as masculine and feminine (terms that imply so much more than gender role instruments assess), I will henceforth call the two factors emotional orientation and instrumental orientation. Both orientations have been linked to health, although each has its own costs and benefits.

NEW RESEARCH DIRECTIONS

According to the work of Myers and Gonda (1982), we all have idiosyncratic concepts of what constitutes a gender role. A promising direction for the field of gender role theory, therefore, would be to measure individuals' definition of masculinity or femininity, and then see how close they feel they are to that concept. Baldwin, Critelli, Stevens, and Russell (1986) constructed a measure of gender role identification (the Sex-Rep) that measures the extent of an individual's identification with his or her personal definition of masculinity or femininity. They noted that the BSRI and Sex-Rep are nonredundant because the latter measures affiliation with personal gender role constructs and the former measures conformity to cultural stereotypes. Use of this new measure may improve the ability of researchers to study the relationship between gender role identification and health, if that is their goal.

A second research direction that seems promising for gender role theorists consists of the examination of sex-differentiating personality traits. It appears that sex-differentiating traits do not form one cluster or category, and indeed factor analyses (e.g., Wilson & Cook, 1984) confirm that personality traits that discriminate between the sexes are multidimensional. It seems reasonable, then, not to study any and all sex-differentiating personality traits and behaviors, but rather to choose as candidates for study those variables that are thought to cause some outcome of interest—in the present case, good health. For example, although applying makeup may be a behavior that differentiates the sexes, it is not a likely candidate for study because it is not regarded a determinant of health. On the other hand, emotional expressivity has been linked to health and also has been found to differentiate the sexes.

It is also important to examine how individuals internalize certain sex-differentiated personality traits; internalization can be considered in two ways. First, Spence (1985) believes that gender identity may be the mechanism through which individuals internalize certain attributes. Gender identity, or the acceptance of one's maleness or femaleness, is established very early in life.

Gender identity . . . is thus a primitive, unarticulated concept of self, initially laid down at an essentially preverbal stage of development and maintained at an unverbalized level. As such, a person's sense of masculinity or femininity is ineffable—in-capable of being put into words. (Spence, 1985, p. 80)

Spence proceeded to explain that while gender identity is being established, individuals may adopt sex-congruent attributes to confirm their sense of gender identity. As individuals age, however, their sense of what constitutes masculine and feminine becomes increasingly complex and is shaped by their particular environment and abilities. Developed in this way, a person's sense of masculinity or femininity is not universally tied to a consistent and limited set of traits or behaviors, but rather to a multidimensional and uniquely constructed constellation of attributes. This would explain why the meanings of masculinity and femininity are so difficult to capture.

An alternative way of understanding how the
The sexes adopt different personality traits in examining the socialization patterns or reinforcement contingencies of boys and girls for various behaviors and personality styles. Although Spence’s gender-identity approach is fascinating, socialization patterns for specific attributes are more easily investigated. (Ultimately, work on Spence’s theory of preverbal gender identity will likely be fruitful, but measurement of such constructs will require some ingenuity.)

Thus, the second direction that gender role theory could move in would be to study specific variables that (a) cause a particular outcome, (b) differentiate men and women, and (c) are differentially reinforced in boys and girls. This direction may provide a useful framework for understanding why modern-day American men report fewer physical and psychological symptoms and die earlier than modern-day American women. In order to elucidate this research direction, three promising personality variables will be examined for evidence of (a) their relationship to health, (b) differentiation between men and women, and (c) different socialization patterns.

**SEX-DIFFERENZIATING VARIABLES AND HEALTH**

**Emotional Orientation**

Both expressing one’s feelings and being aware of one’s emotions have been found to be linked to good health. Although the expression and awareness of emotion are probably related, and together may be considered as a general emotional orientation, they will be discussed separately.

The relationship between emotional expressivity and physiological reactivity has been investigated. Notarius and Levenson (1979), for example, coded the overall facial expressivity of subjects as well as recording heart rates, respiration rates, and galvanic skin responses during a stressful situation. They found that the greater the facial expressive tendency, the less the physiological reactivity (in the form of heart and respiration rate). Thus, those individuals who expressed their emotions during the stressor actually responded less severely on the physiological level than those who expressed less emotion. Additionally, Notarius and Levenson coded expressivity during a second stressful situation and demonstrated that the tendency to inhibit or express emotions was relatively stable.

Although other researchers have not found the inverse relationship between facial expressivity and reactivity (Laird, 1974; Lanzetta, Cartwright-Smith, & Kleck, 1976), many other studies have, even when employing different stressful tasks, different measures of emotional expressivity, and different indices of physiological reactivity (Buck, Miller, & Caul, 1974; Buck, Savin, Miller, & Caul, 1972; Lanzetta & Kleck, 1970).

The inverse relationship found between emotional expressivity and physiological reactivity shows support for the discharge theory of emotion. According to this model, when emotional reactions are aroused they must be discharged, either directly through overt expressions or indirectly through physiological reactivity. There are implications about long-term physical health from the discharge theory of emotion; expressing emotions is thought to be related to decreased physiological reactivity, which in turn is thought to be related to good health. Additionally, inhibiting the expression of emotion will have a deleterious effect on long-term health.

The second facet of an emotional orientation, the experience of emotion, may be associated with decreased psychological health. Watson and Clark (1984), in an extensive, integrative review of past studies, suggested that the disposition to experience and report negative emotion is associated with poor psychological health, even in the absence of overt stress.

Watson and Clark (1984) reviewed many diverse personality scales and demonstrated that a large number of popular tests alleging to measure anxiety, depression, neuroticism, and defensive style all are highly intercorrelated (p. 467). They concluded from this analysis that in fact these measures tap a general disposition to experience negative affect (NA). They then reviewed past research findings and reinterpreted the findings with this new, integrative construct in mind.

For example, White and Wilkins (1973) asked subjects to identify the mood expressed in ambiguous slides while false physiological feedback was being given. High NA individuals interpreted the moods of slides faster than low NA subjects, particularly when high arousal physiological feedback was provided. Watson and Clark (1984) deduced from this finding that high NA individuals are more eager to identify their emotions, whereas low NA individuals are, if anything, reluctant to do so.

Thus it appears that expressing one’s emotions
decreases one's physical health risks, but focusing on one's emotions increases one's psychological health risks. Before we accept this puzzling conclusion, however, we must examine some crucial distinctions between the two sets of results.

One difference between the two general areas of emotion research is that Notarius and Levenson (1979) measured overall emotion, not positive and negative emotions separately, whereas Watson and Clark (1984) concentrated on negative emotion. The facial expressivity in Notarius and Levenson's study, however, was measured during exposure to an industrial accident film and threat of shock. Consequently, it is unlikely that many of the expressions could have been positive. Therefore, the data in both these studies largely consists of negative emotion. Further studies will be needed if we are to examine the effects both of focusing on and of expressing positive affect.

It is also important to note that Notarius and Levenson measured expressivity in response to a stressful task, whereas Watson and Clark addressed a more pervasive tendency to focus on and experience negative affect. The expression of negative emotion in stressful circumstances may be more of a coping style, whereas constant rumination about negative emotion may be more of a pervasive trait or personality style.

The next step is to examine the evidence that men and women differ in their experience and expression of emotion. Allen and Haccoun (1976), while examining variables that influence emotional expressivity, found that women express more emotions than men—particularly fear and sadness. From their other findings, the authors speculate that for females, emotion—particularly fear and sadness—may have a relatively important communicative function: eliciting help.

Brody (1985) also noted pervasive sex differences in emotional expression. She reviewed developmental research on sex differences in expressivity and reported that as boys grow older they decrease in their overall expressiveness, whereas girls decrease in the expression of only a few emotions, notably anger. She concluded that males and females are subject to different social pressures and sanctions concerning the expression and experience of emotions. She also has noted that females may be less subject to social constraints about emotional expression in experimental situations—that is, under stress—than are men.

Although we could speculate that a person must be aware of or experience negative emotions before expressing them—that one cannot have the benefits of the coping style without the costs of the personality style—Watson and Clark (1984) provide evidence that there are individuals who can be aware of and express their negative emotions when stressed, though they are not constantly aware of negative emotion in the absence of stress.

Unfortunately, it may be that individuals currently are socialized to have either both attributes—a constant awareness and an expression of negative emotion—or neither. Thus, it may be that society currently socializes females to focus on their negative emotions generally (i.e., as a personality style) and to express their negative emotions as a coping response, to elicit help. Men, on the other hand, may be socialized to not dwell on their negative emotions generally and also to not express them in response to stress.

If further investigation should support the above conjectures, then emotionality may turn out to be useful in explaining why women differ from men in terms of psychological and physical distress. Constant awareness of negative emotion as a personality style may have deleterious effects on one's psychological health, whereas expression of negative emotion as a response to stress may be beneficial to one's physical health; clearly this sex-differentiating personality variable warrants closer scrutiny.

**Locus of Control**

Locus of control—the attribution of control over events to oneself (internal locus) or to focus outside oneself (external locus)—also has been linked to both physical and psychological health. For example, Cohen and Edwards (1989) reviewed numerous personality characteristics that may act as buffers against stress and concluded that an internal locus of control is one of the few clearly demonstrated stress buffers for psychological health. Consistent with this result, in a longitudinal study Reid and Ziegler (1981) found that internals (in this case, those individuals who had a sense of personal control over the occurrence of desired events) had better physical health and physical ability to do everyday tasks than externals 5 years later. The relationship between locus of control and health is not simple; variables such as level of social support may affect the type of relationship that is found (Lefcourt, Martin, & Saleh, 1984).
In addition, men and women sometimes appear to differ in their levels of perceived control. Doherty and Baldwin (1985) examined longitudinal data from four large national probability samples and found that during the 1960s there were not any large sex differences in locus of control. However, by the mid-1970s women had become more external in their control orientation, whereas men's orientations remained the same. Pearlman and Schooler (1978) also found that women had less of a sense of mastery than men and suggested that the source of this difference may be different socialization patterns.

In their classic text on sex differences, Maccoby and Jacklin (1974) found a consistent sex difference in child-rearing practices: the actions of boys more often had greater consequences than did the actions of girls. That is, both rewards and punishments were contingent upon boys' behaviors. Girls, on the other hand, were less likely to suffer consequences, and so may have had less opportunity to develop a sense of the connection between their behavior and subsequent events.

In sum, society may reinforce women in such a way that they develop an external locus of control, and this personality variable in turn may decrease both their physical and their psychological health.

**Problem Recognition**

The link between problem recognition and health outcomes has been investigated and debated extensively. Often the debate has focused on accurate problem recognition, or stress appraisal. Lazarus (1983) indicated that the question posed usually receives an either/or answer; either an accurate appraisal is said to be a mark of mental health, or it is not. In fact, Lazarus suggested, whether accurate reality testing is costly or beneficial may depend on the context in which it occurs.

One conclusion Lazarus drew from the research he reviewed was that failure to recognize that there is a health problem (i.e., denial) when direct action could change the outcome leads to poor health outcomes. For example, denying that one is having a heart attack is likely to lead to more severe complications than would awareness and recognition of the attack.

Evidence to support Lazarus' speculation comes from Weinstein (1982), who asked students how interested they were in doing more to reduce their own risk for each of 41 health problems. Students also were asked to rate their own worry about each health problem and their perceived risk for each problem compared with other, similar students. The health problems ranged from minor illnesses such as headaches to life-threatening ones such as cancer. The results suggested that unrealistic optimism (i.e., consistently rating one's chance of experiencing a health problem as below average) was related to less worry, and where there is less worry, there is less interest in performing actions to reduce health risks.

On the other hand, when direct action cannot change the outcome, denial should not be destructive and may even have some psychological value. Lazarus (1983) noted that some types of denial (namely partial denial, or illusions) may in fact be beneficial for one's mental health. In reviewing research on this topic, Taylor and Brown (1988) found that some illusions about one's control and self-worth may indeed promote psychological health.

Lazarus (1983) also pointed out that a time perspective may be necessary in order to accurately weigh the costs and benefits of denial. Denial may very well be adaptive during the early stages of a stressful encounter but maladaptive at the later stages. Hofer, Wolff, Friedman, and Mason (1972) studied the grief process among parents of children who were dying of leukemia. At an earlier contact, levels of corticosteroids—which indicate level of physiological distress—were measured for the parents. Initially, denying parents had lower levels of corticosteroids (less physiological distress) than did nondenying parents, who had elevated levels. Many months later, this pattern reversed, and the denying parents had higher levels of corticosteroids than did the nondenying parents. Although this finding has not yet been replicated, it implies that denial may help one through the crisis, but also may have (physical) costs down the road.

Do men and women differ in their problem recognition or stress appraisal? Shaw (1982) found (unexpectedly) that college men reported significantly fewer stressful life events than did college women. Folkman and Lazarus (1980) found that men and women in the larger community reported different sources of stress. Specifically, women reported significantly more health-related stressful episodes than did men. It also may be that the difference found by Shaw (1982) reflects a difference in reporting health-related events, but the type of stressful life event was not reported in his study.

Kessler, Brown, and Broman (1981) analyzed
sex differences in psychiatric help-seeking in four large surveys. Again, women were more likely to have reported a serious life problem (concerning personal, emotional, behavioral, or mental problems) within the last year than were men. Consistent with other epidemiological surveys, they found that women were more likely to have sought psychiatric help. They also found, however, that women were more ready than men to translate a given level of nonspecific distress into a conscious recognition that there was a problem. Once a problem was recognized, men and women were equally likely to seek professional help. Thus, it was a difference in the interpretation of distress as a sign of an emotional problem (i.e., problem recognition) that accounted for the sex difference in seeking psychiatric help.

Lazarus and DeLongis (1983) pointed out that the impact of some stressors "may be obscured by culturally imposed values and constraints and hence be considered unimportant or even go unacknowledged by the person experiencing the stressors" (p. 247). One could hypothesize, then, that men are socialized differently from women in terms of general problem recognition, and that recognition of health problems may be particularly different for men and women. Mechanic (1964) has asserted that boys and girls are taught to interpret physical symptoms differently and to react differently when feeling ill. Also, sex differences in physical-symptom acknowledgment and response increase as children get older (Mechanic, 1964).

Thus, an intriguing picture about sex differences, health-problem acknowledgment, and health begins to emerge. There is some evidence that problem recognition is related to health—positively for controllable stressors and negatively for uncontrollable stressors (although further longitudinal studies may suggest otherwise). Health problems are quite often (although not always) controllable, and it is with health problems that sex differences emerge most clearly. Finally, there is some evidence that girls are taught to recognize and respond to physical and psychological problems differently from boys. So it may be that women, by recognizing health problems more readily than men, take more appropriate health-producing action in the face of threats to health.

**SUMMARY**

Five major hypotheses about the relationship between gender roles and health have been advanced. First, the traditional hypothesis was that sex-appropriate gender roles promote the best health. Second, the balance theory of androgyny predicted that those who possess equal numbers of masculine and feminine attributes will be the healthiest. Third, the main effects androgyny hypothesis suggested that those individuals high in both masculine and feminine attributes will have the additive beneficial effects of both gender roles and so be better off than those who have access to the attributes of only one gender role. Fourth, the emergent properties theory of androgyny suggested that individuals who score high on both the masculine and feminine scales will be in the position to cope with stress most adaptively because they have both a larger coping repertoire and the flexibility to select appropriate coping strategies. Finally, the masculinity hypothesis proposed that the more masculine one is, the healthier one will be. Although there are findings to support each theory, the masculinity hypothesis has garnered the most support across a number of health indexes.

However, limitations to both the research and the theory were noted. Of most importance is the lack of construct validity for the most popular gender role measure, the BSRI. Factor analyses suggest that the masculinity and femininity scales assess, at best, instrumental and emotional orientations, respectively. It is difficult to continue within a theoretical framework when the scales available do not measure what is expected.

Two new research directions were discussed that could help to advance the investigation of sex differences in psychological and physical health. First, one could collect idiographic gender role definitions, as well as subject's conformity with these personally defined gender role definitions. Subsequently, one could examine the relationship between gender role conformity and health. Gender role researchers believed that they were assessing gender role conformity when measuring deviations from gender role stereotypes. Recent evidence, however, suggests otherwise.

When considering gender role attributes that differentiate the sexes, not every possible attribute should be evaluated. Rather, gender role researchers should study specific variables that (a) cause a particular health outcome, (b) differentiate men and women, and (c) are differentially reinforced in boys and girls.

Three personality variables and their relationship to health were examined: emotional orientation, locus of control, and problem recognition. Although it is tempting to equate an emotional
orientation with the outmoded femininity scale, and locus of control with the outmoded masculinity scale, this would be unwise, given the factorial complexity of the two gender role scales. It is probably easier to simply measure the variables that appear to be related to health and then study how (and why) they differ between the sexes.

Finally, it should be noted that understanding sex differences in problem recognition may be pivotal for understanding sex differences in morbidity and mortality. It may be that women are socialized to recognize and respond to health issues more readily than men do, and that this accounts for both the increased health complaints and the longer lives of American women.

REFERENCES


