ORIGINAL ARTICLE

Direct and Indirect Assessment of Gender Role Identification

Sonja van Well · Annemarie M. Kolk · Nicole Y. L. Oei

Abstract Differences in gender role identification exist among both men and women. Earlier researchers have developed several instruments to measure the degree to which individuals identify with the masculine or feminine gender role. In the present study we examined a number of these measurement procedures. Undergraduate students (*N*=45) were administered three direct and two indirect measures of gender role identification. In addition, participants were exposed to a psychological stress test that was relatively masculine. Findings reveal that direct and indirect instruments tap different underlying constructs of gender role identification that are nevertheless positively correlated. Furthermore, results suggest that one of the indirect measures, the Gender Implicit Association Test (GIAT), is a promising instrument to provide an estimate of gender role identification. Of all gender role identification measures the GIAT was (a) most sensitive to sex differences and (b) the only significant predictor of systolic blood pressure responses during and after the relatively masculine stress task.

Keywords Gender role identification · Direct and indirect assessment · Implicit association test · Cardiovascular responses

There is ongoing research on gender role identification and behavior in different domains, for example, in the domain of gendered behavior and vulnerability to health problems.

S. v. Well (✉) · A. M. Kolk · N. Y. L. Oei
Department of Clinical Psychology, University of Amsterdam,
Roetersstraat 15, 1018 WB, Amsterdam, The Netherlands
e-mail: s.m.vanwell@uva.nl

In order to measure gender role adherence, early gender researchers designed gender role instruments from a trait perspective. These measures estimate the degree to which individuals adopt cultural expectations about gender roles to form their self-concept. Frequently used instruments are the Bem Sex Role Inventory (BSRI; Bem 1974) and the Personal Attributes Questionnaire (PAQ; Spence et al. 1974). Both instruments include separate one-dimensional scales of masculinity and femininity and measure gender role identification by means of self-ascribed personality characteristics that are considered more desirable (BSRI) or more stereotypical (PAQ) for one sex than for the other.

Another approach to the assessment of gender role identification is the amount of stress that results from the perceived failure to meet the traditional gender role standards (Good et al. 2000). Eisler and colleagues (Eisler 1995; Eisler and Skidmore 1987) introduced the concept of gender role stress. As they pointed out, gender role stress refers to the tendency to experience stress when faced with behavior, thoughts, or environmental events that challenge one's gender role. To estimate the degree to which certain gender role-related situations are stressful to individuals, Eisler and colleagues developed the Masculine and Feminine Gender Role Stress scales (MGRS and FGRS, respectively; Eisler and Skidmore 1987; Gillespie and Eisler 1992). These Gender Role Stress (GRS) scales contain items that describe situations in which individuals fail to meet the standards of the masculine or feminine gender role. As Good et al. (2000) have argued, the GRS scales can also be used as an alternative approach to measure one's gender role adherence. That is, gender role identification is assessed by the estimation of how stressful it is to deviate from a particular gender role.

The measures described thus far all seek to provide an estimate of gender role identification by asking for a self-report. As with all self-report measures, social desirability

confounds are probable because these reports are influenced by the self-representational goals of participants (Edwards 1957). Consequently, self-report measurement procedures are likely to be distorted to get to the real construct of interest. In general, self-report measures are classified as direct measurement procedures. The GRS questionnaire, however, conceals the assessment of gender role adherence by asking how stressful certain situations would be to individuals rather than asking for self-ascribed masculine and feminine characteristics (as in the BSRI and PAQ). Therefore, that instrument is considered less direct than the BSRI and PAQ. Measurement procedures that assess the construct of interest without requiring individuals to self-assess the extent to which they hold the particular construct are considered indirect (de Houwer 2006). Those measures are less vulnerable to self-representational biases.

A recently developed method to measure gender role adherence indirectly, based on the general-purpose Implicit Association Test (IAT) procedure (Greenwald et al. 1998) is the Gender Implicit Association Test (GIAT; Aidman and Carroll 2003; Greenwald and Farnham 2000). The IAT is a computerized classification task that assesses automatic association strengths between concepts by calculating response latencies. The GIAT assesses the automatic association strength between the self-relevant concept Me (relative to Not me) and the concepts of Masculine and Feminine. It assumes that stronger automatic associations should lead to faster congruent and slower incongruent response latencies. Therefore, individuals with a feminine gender role identification should respond substantially faster when the concepts Me and Feminine share one response key (congruent) than when the concepts Me and Masculine share one response key (incongruent), whereas the reverse should hold for individuals who adhere to the masculine gender role. Androgynous individuals, in addition, should demonstrate equal, relatively low, response latencies. The GIAT thus provides an indirectly measured estimate of gender role identification by considering how strongly the concepts Me and Masculine or Feminine are associated.

Despite the deemed indirectness of the IAT procedure (Greenwald et al. 1998), a recent study has shown that the IAT, although to a limited degree and much less than direct report measures, is also susceptible to faking (Steffens 2004). It can be argued that the IAT measures automatic self-ascription of a certain trait in a relatively direct manner, as the concepts of interest are exposed on the computer screen. Therefore, we developed another indirect method of estimating one's gender role adherence, that is, a gender priming procedure. The underlying mechanism of priming depends on activation spreading. Research on, for example, automatic attitude activation has shown that, when primed with an attitude-object, people can more quickly identify target words that are affectively consistent with the prime

than they can identify target words that are affectively inconsistent (Bargh et al. 1992; Fazio et al. 1986). Based on this activation effect, we argue that gender role identification can be measured indirectly by priming people with a self-relevant stimulus and then assessing the ease with which they can identify masculine versus feminine related target words. This gender priming task (GPT) is presented as a differentiation task between person and object targets, one-half of which had a feminine and one-half of which had a masculine connotation. For masculine participants, the activation of the "self" in response to a self-relevant prime should facilitate the response on the masculine-related targets more than the response on the feminine-related targets, whereas the opposite would hold for participants who adhere to the feminine gender role. The major appeal of this GPT is that there is no need to ask participants for a self-report, or to reveal the concept of interest.

Given the variety of instruments used to provide an estimation of one's gender role identification, it seems useful to investigate these different measures. In a previous study, Greenwald and Farnham (2000) examined sex differences on, and construct divergence between, three gender role identification measures (i.e., the direct BSRI and PAQ and the indirect GIAT). Results of that study showed a large sex difference on the indirect measure along with small to moderate sex differences on the direct measures. Greenwald and Farnham postulated that the difference in effect sizes might be explained by a recent shift in the American view of ideal women and men by which ideal gender roles are more and more overlapping and the fact that direct measures are sensitive to societal pressures, whereas indirect measures are considered to be free of such pressures. The results of the Greenwald and Farnham's (2000) study further revealed construct divergence between the direct and indirect measures. There is a growing body of evidence for construct divergence between direct and indirect measures in several domains, which indicates that these two measurement procedures provide somewhat different information, that is, controlled versus automatic constructs, respectively (Cunningham et al. 2001; Fazio and Olson 2003; Hofmann et al. 2005).

The aim of the present study was to replicate and expand the findings of the Greenwald and Farnham's (2000) study. We included not only the instruments examined by Greenwald and Farnham (i.e., BSRI, PAQ, and GIAT), but also the GRS scales and a gender priming task. The results will allow us to gain insight into the measurement of gender role adherence and can be used in research on gender role identification and behavior in different domains, for example, in the domain of gendered health behavior and vulnerability to health problems.

We, first, examined differences between men and women on the gender role identification measures. As sex and

gender role identification do overlap, sex differences on all instruments are expected. That is, women are more likely to identify with feminine than with masculine traits and behaviors, whereas men are more likely to identify with masculine than with feminine traits and behaviors. Furthermore, based on the results of Greenwald and Farnham (2000), greater sex differences were expected on the indirect than on the direct gender role identification measures.

Then, we investigated whether the direct and indirect measures of gender role identification tap the same, or different, underlying constructs of gender role identification. It was expected that the BSRI, PAQ, and GRS scales (direct measures) would constitute one construct, whereas the GIAT and GPT (indirect measures) would constitute another. To complicate matters the classification of the GRS scales as a direct instrument is debatable. Though the GRS scales require introspection (direct measurement procedure), the questionnaire disguises the assessment of gender role adherence. But does this make the GRS questionnaire an indirect measure? Confirmatory factor analyses were expected to shed more light on the appropriateness of classifying the GRS questionnaire as a direct or more indirect instrument.

An additional aim of the present study was to evaluate the different gender role identification measures as predictors of cardiovascular responses on a psychological stressor that is relatively masculine. It has been suggested that cardiovascular reactivity (CVR) is, in part, a function of the interaction between one's gender role identification and the gender relevance of a stressor. According to this model, individuals who strongly adhere to a gender role show greater CVR to stressors relevant to their gender than to stressors relevant to the other gender or to gender-neutral stressors (Kolk and van Well 2007; Lash et al. 1990; Martz et al. 1995). We expected that the more strongly individuals adhere to the masculine gender role (and the less strongly to the feminine gender role), the higher their CVR on the relatively masculine stressor. We likewise expected that the less strongly individuals adhere to the masculine gender role (and the more strongly to the feminine gender role), the lower their CVR on the relatively masculine stressor. Further, it was hypothesized that the appraisal of a stressor as relevant and the subsequent heightened physiological responses are automatic processes. Indirectly measured attitudes were expected to be more predictive than directly measured attitudes of automatic behavior, whereas directly measured attitudes were expected to be better predictors of controlled behavior than indirectly measured attitudes (Fazio 1990; Perugini 2005). Asendorpf et al. (2002), among others, were able to demonstrate such a double dissociation model for trait shyness and shy behavior. That is, indirectly assessed shyness (IAT) uniquely predicted spontaneous shy behavior, whereas directly measured

shyness (self-ratings) uniquely predicted controlled shy behavior. Based on these results, it was expected that the indirect gender role identification measures would be better predictors of cardiovascular responses (automatic behavior) than the direct instruments would.

Method

Participants

Participants were recruited by means of a sign-up board posted at the University of Amsterdam. Eligibility criteria included no hypertension (i.e., blood pressure not higher than 140/90 mmHg), no history of cardiovascular disease, no chronic disease that requires medical attention, no current use of prescribed medication, and a body mass index (BMI; kg/m^2) between 19 and 25. Twelve respondents were excluded from participation because they did not meet all eligibility criteria ($n=5$) or because they had lost interest in participating ($n=7$). The final sample consisted of 22 female and 23 male undergraduate psychology students, aged between 17 and 36 years ($M=21.0$, $SD=3.2$). Each participant gave signed informed consent in which confidentiality, anonymity, and the opportunity to withdraw without penalty were assured. Participants received course credit for taking part in the study.

Direct Measures

Bem Sex Role Inventory (BSRI; Bem 1974)

The BSRI consists of 60 characteristics. Twenty characteristics, which are considered more desirable for men than for women, represent the masculinity scale (e.g., independent), whereas 20 characteristics, which are considered more desirable for women than for men, represent the femininity scale (e.g., tender). The remaining items serve as filler items. Participants rate how well each characteristic applies to them on a 7-point Likert scale that ranges from 1 (*never or almost never true*) to 7 (*always or almost always true*). Both scales have been found to be reliable and valid (Bem 1974; Holt and Ellis 1998). In the present sample, Cronbach's alpha was .83 and .79, for the masculinity and femininity scale, respectively. The BSRI was translated into Dutch according to back-translation rules (Brislin 1986).

Personal Attributes Questionnaire (PAQ; Spence et al. 1974)

The PAQ consists of 24 trait dimensions. The masculinity scale contains eight items stereotypically more associated with men than with women (e.g., self-confident), whereas

the femininity scale contains eight items stereotypically more associated with women than with men (e.g., understanding of others). The remaining items serve as filler items. Participants rate each item as to how much it applies to them on a 5-point scale with the endpoints labeled with opposites (e.g., very self-confident, not at all self-confident). Validity and reliability of the PAQ have been found to be satisfactory (Helmreich et al. 1981). Based on present data, Cronbach's alpha was .75 and .69, for the masculinity and femininity scale, respectively. The PAQ was translated into Dutch according to back-translation rules (Brislin 1986).

Gender Role Stress (GRS; Eisler and Skidmore 1987; Gillespie and Eisler 1992; Dutch Translation by van Well et al. 2005)

The Masculine Gender Role Stress (MGRS) scale contains 40 items that describe situations that elicit stress in relation to the perceived failure to meet the standards of the masculine gender role (e.g., appearing to be less athletic than a friend), whereas the Feminine Gender Role Stress (FGRS) scale consists of 39 items that refer to failure to meet the standards of the feminine gender role (e.g., having someone else raise your children). Participants rate each item on a 6-point Likert scale that ranges from 0 (*not stressful*) to 5 (*extremely stressful*). The reliability and validity of both GRS scales have been found satisfactory (Eisler and Skidmore 1987; Eisler et al. 1988; Gillespie and Eisler 1992). Furthermore, the Dutch version of the GRS scales was found to be highly reliable and cross-culturally valid (van Well et al. 2005). In the current sample Cronbach's alpha was .92 and .94 for the MGRS and FGRS scale, respectively.

Indirect Measures

Gender Implicit Association Test (GIAT; Greenwald and Farnham 2000)

The GIAT was based on the procedure described by Greenwald and Farnham (2000). The GIAT requires participants to use two response keys to categorize words as belonging to one of four categories. Categories and stimuli used in the present GIAT are (a) Me: I, self, me, my, mine; (b) Not me: they, them, it, their, other; (c) Feminine: woman, girl, lady, madam, daughter; and (d) Masculine: man, boy, sir, gentleman, son. After practicing the Me/Not me discrimination and the Feminine/Masculine discrimination separately, the two categorization tasks were combined. This combined task represents the experimental task that was administered twice. First, Me and Feminine categories shared the left response key, and the categories Not me and Masculine shared the right response key.

Second, the categories Me and Masculine were assigned to the left response key, and the Not me and Feminine categories were assigned to the right response key. The discrepancy in response latencies between the two combined blocks represents gender role identification. Administration order of the combined blocks as well as key assignment was counterbalanced. Each experimental block consisted of a block of 20 practice trials followed by a block of 40 experimental trials. To minimize variability in response latencies for the first few experimental trials, each block was preceded by three "warm-up" trials. Warm-up trials were excluded from further analyses.

Target items appeared, one at a time, in the centre of a computer screen in a randomly selected order. Category labels were presented in the upper right and left corners of the screen. After a 500 ms interval a fixation cross was presented for 500 ms followed by the next stimulus. Stimuli disappeared after a response was made or after 5,000 ms. Participants received no accuracy feedback.

Data were treated in accordance with the improved *D* score algorithm recommended by Greenwald et al. (2003). The GIAT effect was calculated based on both practice and experimental trials of the two experimental blocks. For these practice trials and experimental trials, separately: (a) a standard deviation was calculated on correct responses; (b) error trials were replaced with the mean of correct responses plus a 600 ms penalty; (c) response latencies of the two experimental blocks were averaged and the resulting means were subtracted; and (d) the difference score was divided by its matching standard deviation. The resulting *D* scores on the experimental and practice trials were then averaged. The GIAT effect was computed such that higher scores represent stronger masculine gender role identification, whereas lower scores reflect stronger adherence to the feminine gender role.[1]

The validity of several IAT measures has been supported repeatedly, and the reliability has been found satisfactory, with good internal consistency estimates but poorer test-retest reliabilities (Bosson et al. 2000; Fazio and Olson 2003; Greenwald and Nosek 2001). Cronbach's alpha was .88 for the GIAT in the current sample. Cronbach's alpha was computed following procedures described by Bosson et al. (2000), so that it reflects the internal consistency in the tendency to associate feminine, relative to masculine, with the self.

Gender Priming Task (GPT)

The GPT was based on supraliminal attitude-prime tasks used in research on automatic attitude activation (Bosson

---

[1] Our use of this scoring does not imply that femininity or women are less than masculinity or men.

et al. 2000; Fazio et al. 1986). This GPT requires participants to differentiate target items that belong to the category Persons or Objects, after having been exposed to a self-relevant prime (me) or a self-irrelevant prime (they). Each category included five feminine- and five masculine-related target items. Targets used were (a) feminine persons: woman, girl, lady, madam, daughter; (b) masculine persons: man, boy, sir, gentleman, son; (c) feminine objects: handbag, novel, make-up, wine, blow-drier; and (d) masculine objects: football, motorcycle, fishing rod, beer, saw. Key assignment was counterbalanced.

Category labels were presented in the upper right and left corners of the screen throughout the task. Each trial started with a fixation cross in the centre of the computer screen. Thereafter, the prime appeared for 200 ms followed by a 100-ms buffer. Next, a randomly selected target item appeared. Targets disappeared after a response was made or after 5,000 ms. Participants received no accuracy feedback. The intertrial interval was 500 ms. Each target item was paired once with the self-relevant prime and once with the self-irrelevant prime for a total of 40 categorization trials. A 10-trial practice block and three warm-up trials preceded the 40-trial experimental block.

Data on the experimental trials preceded by the self-relevant prime were included in further analyses only. These data were treated in accordance with the procedure described for the GIAT. Response latencies were averaged on the masculine and feminine trials, separately. The resulting means were subtracted such that a higher score represents greater activation spreading between self and masculine relevant trials than between self and feminine relevant trials.

The validity and reliability of various priming measures has been supported, although not consistently (Bosson et al. 2000; Fazio and Olson 2003). Cronbach's alpha of the GPT was calculated following the procedure described for the GIAT. The reliability of the priming task was low; Cronbach's alpha was .36. This low reliability is not unusual in priming procedures (Bosson et al. 2000; Fazio and Olson 2003). Despite its disappointing low reliability the GPT was included in the analyses.

Psychological Stressor

Task

Based on the Trier Social Stress Test (Kirschbaum et al. 1993), the stressor consisted of a 10-min anticipation period followed by a test period in which participants had to deliver a speech for a job application (5 min) and perform a N-Back task (5 min) in front of a three-person selection committee.

Participants were instructed to take the role of job applicant and to imagine that they had been invited to introduce themselves to a selection committee. They were asked to deliver a 5-min free speech in which they tried to convince the committee that they were the best person for the job. They were urged to make a believable impression, because the committee would ask questions in case of incredibility. Furthermore, participants were told that the committee would take notes about the content and manner of the speech. The selection committee was introduced as consisting of two psychologists and one future colleague. In case the participant finished the speech in less than 5 min, a committee member responded in accordance with a standardized protocol.

For the N-Back task a set of 100 randomly generated digits was constructed and presented in a fixed order. Participants were asked to indicate whether each auditory-presented digit was similar to (target) or different from (non-target) the digit presented three digits before by saying out loud "yes" to a target and "no" to a non-target. The task consisted of 30% targets. Participants were instructed to give as many correct answers as possible. One committee member responded to incorrect answers by saying out loud "incorrect," whereas another member marked the participant's performance by means of a scoreboard.

Men have been found to appraise challenges to occupational and intellectual abilities as more stressful than women do (Eisler and Skidmore 1987). As the stress task used in the present study challenges successful performance with regard to work (getting the job, being the best candidate for the job) and successful performance on the N-back, the stressor was defined as relatively masculine. We classified the stressor as *relatively* masculine because we acknowledge that there are other stress tasks of which the masculine relevance is more pronounced (e.g., stressors in which the gender relevance is manipulated; see Kolk and van Well 2007; Lash et al. 1990).

Cardiovascular Measures

Systolic blood pressure (SBP, mmHg), diastolic blood pressure (DBP, mmHg), and heart rate (HR, bpm) were recorded with a Finapres blood pressure monitor (Finapres 2300, Ohmeda, Englewood, CO, USA). The Finapres enables non-invasive continuous beat-to-beat monitoring of the finger arterial pressure waveform using a finger cuff applied to the middle phalanx of the middle finger (see also Imholz et al. 1998).

Measures of cardiac output (CO, l/min) and total peripheral resistance (TPR, $dyn.s/cm^5$) were derived from the Finapres data with BeatScope version 1.1 (TNO-Biomedical Instrumentation, Amsterdam, The Netherlands). BeatScope is a software package for the analysis of arterial pressure waveforms. It provides the computation of hemodynamic measures with the Modelflow method based on the

simulation of a model of aortic input impedance. Good agreement of these parameters has been obtained with intra-arterial measures (Jellema et al. 1996; Wesseling et al. 1993).

Procedure

Screening

A sign-in board posted at our department presented the study as one about stress and emotion. Respondents were informed about the study protocol in a manner that carefully avoided any reference to sex or gender differences, and they were screened for eligibility criteria over the telephone. If they met all criteria, they were invited to participate in the study, and three laboratory sessions were scheduled. Furthermore, respondents were asked to refrain from eating, smoking, and exercising at least 90 min prior to the session that included the physiological recordings and to refrain from caffeine and alcohol at least 8 h prior to that session. In addition, respondents received a letter that reiterated the information provided during the screening.

Laboratory Sessions

Each participant was tested individually between 9:00 A.M. and 12:00 noon on three consecutive days by one of two female experimenters. At the beginning of each session the experimenter informed the participant about the experimental procedure of the current session, again carefully avoiding any reference to sex or gender differences. Furthermore, the experimenter explained that all instructions and tasks would be provided on a computer screen, and showed the participant the appropriate response keys. In addition, the experimenter noted that the session was monitored via an intercom. She then went to an adjacent room and started the computerized protocol (using the VSRRP98 software package developed at our department). Participants received all further instructions, questionnaires, and tasks by means of the computer screen and provided all their responses with the response keys, except for the instruction and responses with regard to the N-back task, which was part of the stressor.

In the first session participants read and signed the informed consent form. Thereafter the direct measures of gender role identification were administered item by item in two blocks in a fixed order (Block 1, BSRI and PAQ; Block 2, GRS scales). The BSRI and PAQ were presented as personality inventories, whereas the GRS scales were combined and presented as one questionnaire set that dealt with the topic of stress experience. Completion took about 30 min.

In the second session, the indirect measures of gender role identification were administered (using the WESP software package developed at our department). The indirect instruments were presented as reaction time tasks to measure categorization speed. Participants were instructed to complete the tasks as quickly and accurately as possible. Administration order of the measures was counterbalanced. Completion took 30 min on average.

During the third session the experimenter first checked the criteria pertaining to food, cigarettes, caffeine, alcohol, and exercise. All participants met these criteria, therefore it was not necessary to reschedule any participants. The experimenter then attached an appropriate-size Finapres finger cuff to the mid-phalanx of the third finger of the left hand. The left arm was positioned at heart level; if necessary, towels were used to increase the height comfortably. Participants were instructed to minimize all movement during the physiological recordings. After adaptation to the Finapres, a 15-min baseline period followed in which participants were asked to rest quietly while watching a documentary about Tibet in order to get proper physiological baseline levels. After a 10-min speech preparation period (anticipation), the selection committee entered the room, greeted the participant, and took a seat behind a table. Then the participant delivered the speech and performed the N-back task. Thereafter, the selection committee thanked the participant for his/her cooperation and left the room. A 10-min recovery period followed in which participants were asked to rest quietly while watching the second segment of the documentary about Tibet. Subsequently the experimenter removed the finger cuff, and participants completed an exit questionnaire. At the end of the session participants received their course credit. The third session took about 60 min. The order of sessions two and three was counterbalanced.

Analyses

Independent sample t-tests were performed on the scores of all gender role identification measures to determine sex differences. Furthermore, to investigate the association between the direct and indirect measures, Pearson product-moment correlation coefficients were calculated. To examine whether the direct and indirect measures actually tap different constructs of gender role adherence, confirmatory factor analyses (CFAs) were performed. The CFAs were based on a covariation matrix and conducted with LISREL 8 (Jöreskog and Sörbom 1993). To evaluate model fit, Hu and Bentler (1999) recommended a two-index presentation strategy using the standardized root mean squared residual (SRMR) in conjunction with, for example, the comparative fit index (CFI) or the root mean squared error of approximation (RMSEA). For a relatively good model fit, SRMR values below .08, CFI values above .95, and RMSEA values below .06 are required. Furthermore, a

Table 1 Means, standard deviations, and results of *t*-tests for sex differences on direct and indirect gender role identification measures.

	Men (n=22)		Women (n=22)		Difference Men–Women	
	M	SD	M	SD	T	Cohen's d
Direct measures						
BSRI femininity	4.65	.52	4.90	.44	−1.72*	.52
BSRI masculinity	4.63	.45	4.25	.64	2.22*	.67
PAQ femininity	3.70	.50	3.88	.34	−1.33	.40
PAQ masculinity	3.71	.41	3.46	.58	1.66	.50
FGRS	2.92	.54	3.65	.71	−3.86**	1.16
MGRS	2.57	.53	2.60	.58	−.22	.07
Indirect measures[a]						
GIAT	.36	.39	−.74	.31	10.42**	3.14
GPT	.38	.75	−.20	.56	2.88*	.87

BSRI = Bem Sex Role Inventory; PAQ = Personal Attributes Questionnaire; MGRS = Masculine Gender Role Stress; FGRS = Feminine Gender Role Stress; GIAT = Gender Implicit Association Test; GPT = Gender Priming Task
[a] Higher values indicate greater identification with masculinity.
*p<.05, one-tailed
**p<.001, one-tailed

stress manipulation check was conducted on all cardiovascular measures with a repeated measures multivariate analysis of variance (MANOVA) with stress phase (baseline, anticipation, stress, recovery) as the within-subjects factor. Finally, simultaneous multiple regression analyses were conducted to test the predictions regarding the gender role identification measures and their relationship with cardiovascular responses on a relatively masculine stressor. Cardiovascular responses during anticipation, stress, and recovery were computed as difference scores from baseline.

Results

Missing values (<.5%, observed only on direct measures) were replaced with corrected item mean substitutions (Huisman 1999). Data from one male participant were removed because that person was an outlier on one of the gender role identification measures (*z*-score MGRS>3). In addition, due to technical problems with the Finapres, data from another male participant were excluded from the regression analyses only.

Initial analyses revealed no significant effect for (a) session order; (b) administration order of indirect tasks; (c) key assignment in indirect tasks; or (d) order of experimental blocks in the GIAT. Therefore, these variables were dropped in subsequent analyses.

Sex Differences

Table 1 shows means and standard deviations of all gender role identification measures, for male and female partic-

ipants separately. Furthermore, Table 1 presents the results of independent *t*-tests for sex differences and Cohen's d effect sizes (Cohen 1977). As can be seen, five of eight gender role identification measures revealed a significant sex difference in the predicted direction. That is, women more strongly associated the self with femininity than with masculinity, whereas men more strongly associated the self with masculinity than with femininity. Although the PAQ femininity and masculinity scale revealed a similar sex difference pattern, female and male students did not differ significantly on these measures. As for effect sizes, according to Cohen's (1977) convention, results on the GIAT showed a very large effect, followed by large effects

Table 2 Correlations among direct and indirect gender role identification measures and sex (N=44).

|	1	2	3	4	5	6
1. Sex[a]	–	.42**	.33*	.65**	.85**	.41**
Direct measures[b]						
2. BSRI bipolar		–	.75**	.48**	.37*	−.15
3. PAQ bipolar			–	.42**	.28	−.06
4. GRS bipolar				–	.65**	.14
Indirect measures[b]						
5. GIAT					–	.48**
6. GPT						–

BSRI = Bem Sex Role Inventory; PAQ = Personal Attributes Questionnaire; GRS = Gender Role Stress; GIAT = Gender Implicit Association Test; GPT = Gender Priming Task
[a] Women were coded as 1, men as 2.
[b] Higher values indicate greater identification with masculinity.
*p<.05, two-tailed
**p<.01, two-tailed
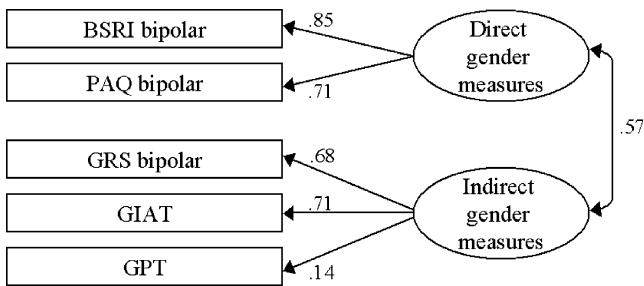
Fig. 1 Factor structure of direct and indirect gender role identification measures. *BSRI* = Bem Sex Role Inventory; *PAQ* = Personal Attributes Questionnaire; *GRS* = Gender Role Stress; *GIAT* = Gender Implicit Association Test; *GPT* = Gender Priming Task.

on the FGRS scale and the GPT, and medium to large effects on the BSRI femininity and masculinity scales. Moreover, no sex differences were found on the MGRS scale.

Construct Divergence

Correlation coefficients among all direct and indirect gender role identification measures and sex are reported in Table 2. Bipolar scores were used for the direct measures (by subtracting the femininity score from the masculinity score). The direct measures were positively and highly intercorrelated, average $r=.55$, all $ps<.01$, as were the indirect measures, $r=.48$, $p<.01$. Furthermore, the GIAT positively correlated with the BSRI and GRS, $r=.37$, $p<.05$ and $r=.65$, $p<.01$, respectively. The remaining correlations between the direct and indirect measures were not significant, which suggests construct divergence between the direct and indirect measures of gender role identification.

To investigate whether the direct and indirect measures actually tap different constructs of gender role adherence, confirmatory factor analyses (CFAs) were conducted. First, a one-factor model, in which all measures represent one underlying gender role identification construct, was examined. Following Hu and Bentler's (1999) two-index presentation strategy, the fit indices of the one-factor model indicated

lack of fit, SRMR=.13; CFI=.76; RMSEA=.28. Then, a two-factor model was evaluated in which the direct measures made up one factor, and the indirect measures made up another factor. However, this model showed lack of fit also, SRMR=.12; CFI=.80; RMSEA=.27. In addition, a second two-factor model was examined. In this model the GRS was removed from the factor that contained the direct measures, and added to the indirect measures factor. This alternative two-factor model met Hu and Bentler's (1999) criterion of model fit, SRMR=.04; CFI=.99; RMSEA=.00. Figure 1 shows the alternative two-factor model. This model shows that the BSRI and PAQ scores serve as indicators of a direct measured construct of gender role adherence, whereas the GRS, GIAT, and GPT scores are indicators of an indirect assessed construct of gender role adherence. The different constructs were nevertheless positively correlated, $r=.57$, $p<.01$. Moreover, additional analyses in which the GPT was excluded revealed similar results.

Predicting Cardiovascular Responses

Table 3 presents means and standard errors of the cardiovascular measures by stressor phase. A repeated measures MANOVA revealed a significant main effect of stress phase, $F(15, 28)=55.71$, $p<.001$, which indicates that the stress test was able to produce significant changes on physiological arousal. Pairwise comparisons (adjusted for inflation of alpha) revealed that all cardiovascular measures, except for TPR, increased from baseline to anticipation to the stress test and decreased thereafter (see Table 3). TPR levels did not change from baseline to anticipation, then significantly increased from anticipation to the stress phase, and remained elevated during recovery.

Furthermore, multiple simultaneous regression analyses were conducted to evaluate the direct and indirect gender role identification measures as predictors of cardiovascular responses on a psychological stress test that is relatively masculine. In order to keep the number of predictors small (considering the relative small sample size) the PAQ was

Table 3 Means and standard errors (in parenthesis) of the cardiovascular measures by stressor phase ($N=43$).

Stress phase	Cardiovascular measures				
	SBP (mmHg)	DBP (mmHg)	HR (bpm)	CO (l/min)	TPR (dyn.s/cm^5)
Baseline	127.4 (2.5)$_a$	75.8 (1.8)$_a$	69.3 (1.9)$_a$	5.21 (.22)$_a$	1,527 (65)$_{ab}$
Anticipation	139.6 (3.0)$_b$	81.3 (2.1)$_b$	74.5 (2.2)$_b$	5.72 (.26)$_b$	1,507 (60)$_a$
Stress	176.1 (3.8)$_c$	103.6 (3.0)$_c$	84.9 (2.2)$_c$	6.32 (.28)$_c$	1,795 (124)$_{bc}$
Recovery	150.2 (3.8)$_d$	89.3 (2.9)$_d$	69.2 (1.7)$_a$	5.19 (.22)$_a$	1,860 (121)$_c$

Means in the same column that do not share subscripts differ at $p<.05$ in pairwise comparisons with Bonferroni adjustment for multiple comparisons. *SBP* = Systolic Blood Pressure; *DBP* = Diastolic Blood Pressure; *HR* = Heart Rate; *CO* = Cardiac Output; *TPR* = Total Peripheral Resistance

Table 4 Summary of simultaneous multiple regression analyses for variables predicting SBP during stress and recovery ($N=43$).

Variable[a]	Stress			Recovery		
	B	$SE\ B$	β	B	$SE\ B$	β
BSRI bipolar	2.06	3.93	.09	.70	4.26	.03
GRS bipolar	−9.60	6.27	−.29	−11.44	6.79	−.34
GIAT	17.11	5.79	.63*	15.80	6.27	.56*
GPT	−5.46	4.29	−.23	−1.58	4.64	−.06

$R^2 =.22$ ($p<.05$) and .18 ($p>.05$) for stress and recovery, respectively. SBP = Systolic Blood Pressure; BSRI = Bem Sex Role Inventory; GRS = Gender Role Stress; GIAT = Gender Implicit Association Test; GPT = Gender Priming Task
[a] Higher values indicate greater identification with masculinity.
*$p<.01$, one-tailed

excluded due to its high similarity to the BSRI. The linear combination of the gender role identification measures (BSRI, GRS, GIAT, and GPT) was significantly related to SBP during stress only, $R^2=.22$, $F(4, 38)=2.66$, $p<.05$. Analyses of the unique strength of each gender role identification measure as individual predictor of cardiovascular responses indicated that the GIAT was the only significant predictor of SBP during stress, $\beta=.63$, $p<.01$, as well as during recovery, $\beta=.56$, $p<.01$. These findings suggest that the more men and women adhere to the masculine gender role, as measured by the GIAT, the higher their SBP during the relatively masculine stressor and the slower their SBP recovery from it. Table 4 shows a summary of the regression analyses of variables predicting SBP during stress and recovery.

Moreover, additional regression analyses that controlled for appropriate baseline level as well as biological sex revealed unchanged findings. Likewise, excluding the priming task from the regression analyses did not alter the magnitude or pattern of observed findings.

Discussion

In the present study we examined several direct and indirect measures of gender role identification. The reported findings suggest that one of the indirect measures, the Gender Implicit Association Test, is a promising tool to provide an estimate of gender role identification and can be adopted in research on gender role adherence. The results demonstrated that, of all gender role identification measures, the GIAT was the most sensitive to sex differences in gender role identification. Results also revealed that the GIAT was the only significant predictor of SBP reactivity and recovery. Furthermore, the present findings showed that, after we reclassified one direct instrument as a more indirect one, direct and indirect gender role identification instruments tap different underlying constructs of gender role identification that are nevertheless positively correlated.

The hypothesis that sex differences on all gender role identification measures would be found was partly supported. Whereas the MGRS scale revealed no sex difference, sex differences were found in the expected direction on the remaining gender role identification measures. That is, although the difference was not significant for the PAQ femininity and masculinity scale, women showed more identification with femininity than with masculinity, whereas men showed more identification with masculinity than with femininity. Furthermore, in agreement with the results of Greenwald and Farnham (2000), the current results generally reveal larger effect sizes for sex differences on the indirect measures than on the direct measures. Nevertheless, it should be noted that, of the two different indirect measures, the GIAT is by far the stronger instrument. According to Greenwald and Farnham (2000), the finding that indirect measures of gender role identification reveal larger sex differences than do direct measures could be explained by the fact that gender roles are more and more overlapping and the fact that direct measures are more sensitive to these societal pressures than are indirect measures. The finding that the BSRI and PAQ revealed smaller sex differences might also be explained by the fact that both measures are over 30 years old. It could be argued that the traits that were considered masculine or feminine at that time do not apply as well today. In this light, it is interesting that, for both instruments, reliability coefficients were lower on the femininity than on the masculinity scales. This might reflect changing norms in particular on the traits that were once considered feminine (e.g., gullible, does not use harsh language).

Consistent with the idea that direct measures tap different information than indirect measures, most of the correlations between the direct and indirect measures were not significant. However, initial CFAs rejected the two-factor model in which the direct self-report measures, on the one hand, and the automatically estimated indirect gender role identification measures, on the other hand, represent two separate constructs. Nevertheless, additional

CFAs supported an alternative two-factor model in which the BSRI and PAQ make up one construct, whereas the GRS scales, GIAT, and GPT constitute another. The shift of the GRS scales from a direct to an indirect measure of gender role identification is explicable. Although the GRS scale is a self-report questionnaire, it does not measure gender role identification directly in terms of self-ascribed masculine and feminine personality traits, but more indirectly, in terms of how stressful it is to deviate from the gender role standards involved. Moreover, the different constructs of gender role adherence are positively correlated. This substantial correlation might be due to the inclusion of the GRS scale to the indirect measures factors, which render this factor somewhat more direct. Taken together, the correlations and CFAs results support the hypothesis that direct and indirect gender role identification measures assess distinct constructs of gender role identification, and these findings are in line with other researchers' findings that reveal construct divergence between direct and indirect measurement procedures (Bosson et al. 2000; Fazio and Olson 2003; Greenwald and Farnham 2000).

Furthermore, the utility of the gender role identification measures as unique predictors of cardiovascular responses was supported for the GIAT on SBP responses only. After we controlled for the other gender identification scores, GIAT scores indicative of a masculine gender role adherence predicted higher SBP reactivity and slower SBP recovery in relation to the relatively masculine stressor. This finding is in line with research that has shown that CVR is a function of the interaction between one's gender role identification and the gender relevance of a stressor (Lash et al. 1990; Martz et al. 1995).

The finding that the GIAT was a better predictor of cardiovascular reactivity and recovery (at least on SBP) compared to the direct measures is in line with Fazio's (1990) double dissociation model. This model implies that indirect measures are better predictors of automatic behavior, whereas direct measures better predict controlled behavior. Unfortunately, the results of the current study support a "single" dissociation between the direct measures of gender role identification and the automatic physiological stress responses only. To be able to test the full double dissociation model, future researchers should also obtain controlled behavioral responses (e.g., self-reported stress ratings) to determine whether indirect gender role identification measures are dissociated from controlled behavior.

However, the other indirect measures, that is, the GPT and (based on CFAs) the GRS scales, were not significant predictors of any of the physiological stress responses. The finding that the GRS scales, unlike the GIAT, had no predictive value can be explained by the difference in the extent to which both measurement procedures are indirect.

Although CFAs demonstrated that the GRS score could be seen as a more indirectly measured estimation of gender role adherence, the GRS scale is a self-report measure. Subsequently, this instrument is more sensitive to deliberate processing than is the GIAT, which makes it less likely to reveal an association between the GRS scale and the physiological stress responses that involve spontaneous processing.

The GPT, on the other hand, represents a more indirect measurement procedure than the GIAT, as during this last procedure the concepts of interest were exposed on the computer screen. Nevertheless, the GIAT, rather than the GPT, predicted cardiovascular responses. This result can be ascribed to the low reliability of the GPT. Therefore, the GPT provided an unstable and potentially inadequate estimate of gender role identification. Only one self-relevant and one self-irrelevant prime were used (me vs. they). Participants might have habituated to these two primes and stopped processing their meaning, thereby diminishing the priming effect. Accordingly, including a larger number of primes might have improved the priming task. This issue needs to be further explored in future research. Moreover, in the current form, the GPT might not be an optimal estimate of gender role identification as we are not sure whether the feminine and masculine persons used as target stimuli (e.g., lady, madam, sir, gentlemen) are gender role adherents. The GPT might be improved by replacing the Persons category and by using more apparent gender role-related target items (e.g., Characteristics as category and sensitive and assertive as feminine- and masculine-related target items, respectively).

Several limitations to our study should be noted. First, the relative small sample size may have lowered statistical power and limited proper interpretation of the results. Second, the use of a student population lowered the external validity and could have reduced the results, as the strength of gender role identification is less strong in a student population than in a more heterogeneous group of participants. Third, a *relatively* masculine stress task was used. Within the academic context and given that the participants were students, the performance-related stressor could also be perceived as gender-neutral in nature. A more pronounced masculine-relevant stressor could have revealed stronger and more convincing results. Finally, the present study included only a stressor that was relatively masculine. It would also be useful to demonstrate that GIAT scores indicative of feminine gender role identification predict cardiovascular responses to a feminine-relevant stressor. Different type of stressors (i.e., masculine, feminine, and gender-neutral) would have helped to clarify the interpretation of the present results. To scrutinize the relations between gender role identification, gender relevance of a stressor, and cardiovascular responses, future

researchers should use a stronger gender-relevant stressor and different gender-relevant conditions. The gender relevance of a stressor could, for example, be experimentally manipulated by varying the instruction preceding the stressor (Kolk and van Well 2007; Lash et al. 1990).

In conclusion, the present study revealed that direct and indirect measures tap different constructs of gender role identification. Furthermore, our data show that one of the indirect measures, the GIAT, is a promising tool to provide an estimate of gender role identification. It can be adopted and used in research on gender role adherence and behavior in different domains. With regard to the relationship between gender role identification and cardiovascular responses, of all gender role identification measures examined in this study, only the GIAT predicted SBP reactivity and recovery on a laboratory stressor classified as masculine-relevant. However, the exact relationships between gender role identification, gender relevance of a stressor, and cardiovascular responses await further examination in future research.

References

Aidman, E. V., & Carroll, S. M. (2003). Implicit individual differences: Relationships between implicit self-esteem, gender identity, and gender attitudes. *European Journal of Personality, 17*, 19–36.

Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology, 83*, 380–393.

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology, 62*, 893–912.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42*, 155–162.

Bosson, J. K., Swann, W. B., Jr., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology, 79*, 631–643.

Brislin, R. W. (1986). The wording and translation of research instruments. In W. J. Lonner & J. W. Berry (Eds.), *Field methods in cross cultural research* Vol. 8 (pp. 137–164). Beverly Hills, CA: Sage.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science, 121*, 163–170.

de Houwer, J. (2006). What are implicit measures and why are we using them? In R. W. Wiers & A. W. Stacy (Eds.), *The handbook of implicit cognition and addiction* (pp. 10–28). Thousand Oaks, CA: Sage.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York: Holt, Rinehart, and Winston.

Eisler, R. M. (1995). The relationship between masculine gender role stress and men's health risk: The validation of a construct. In R. F. Levant & W. S. Pollack (Eds.), *A new psychology of men* (pp. 207–225). New York: Basic Books.

Eisler, R. M., & Skidmore, J. R. (1987). Masculine gender role stress: Scale development and component factors in the appraisal of stressful situations. *Behavior Modification, 11*, 123–136.

Eisler, R. M., Skidmore, J. R., & Ward, C. H. (1988). Masculine gender-role stress: Predictor of anger, anxiety, and health-risk behaviors. *Journal of Personality Assessment, 52*, 133–141.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology*, Vol. 23 (pp. 75–109). San Diego: Academic.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*, 297–327.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*, 229–238.

Gillespie, B. L., & Eisler, R. M. (1992). Development of the Feminine Gender Role Stress Scale: A cognitive-behavioral measure of stress, appraisal, and coping for women. *Behavior Modification, 16*, 426–438.

Good, G. E., Sherrod, N. B., & Dillon, M. G. (2000). Masculine gender role stressors and men's health. In R. M. Eisler & M. Hersen (Eds.), *Handbook of gender, culture, and health* (pp. 63–81). Mahwah, NJ: Erlbaum.

Greenwald, A. G., & Farnham, S. D. (2000). Using the implicit association test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology, 79*, 1022–1038.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*, 1464–1480.

Greenwald, A. G., & Nosek, B. A. (2001). Health of the implicit association test at age 3. *Zeitschrift für Experimentelle Psychologie, 48*, 85–93.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197–216.

Helmreich, R. L., Spence, J. T., & Wilhelm, J. A. (1981). A psychometric analysis of the personal attributes questionnaire. *Sex Roles, 7*, 1097–1108.

Hofmann, W., Gawronski, B., Gschwender, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31*, 1369–1385.

Holt, C. L., & Ellis, J. B. (1998). Assessing the current validity of the Bem sex role inventory. *Sex Roles, 39*, 929–941.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Huisman, M. (1999). Simple and effective methods to treat missing item responses. *Kwantitatieve Methoden, 62*, 57–78.

Imholz, B. P., Wieling, W., van Montfrans, G. A., & Wesseling, K. H. (1998). Fifteen years experience with finger arterial pressure monitoring: Assessment of the technology. *Cardiovascular Research, 38*, 605–616.

Jellema, W. T., Imholz, B. P., van Goudoever, J., Wesseling, K. H., & van Lieshout, J. J. (1996). Finger arterial versus intrabrachial pressure and continuous cardiac output during head-up tilt testing in healthy subjects. *Clinical Science, 91*, 193–200.

Jöreskog, K., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language.* Hillsdale, NJ: Erlbaum.

Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The 'Trier Social Stress Test': A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology, 28*, 76–81.

Kolk, A. M., & van Well, S. (2007). Cardiovascular responses across stressor phases: The match of gender and gender-role identification with the gender relevance of the stressor. *Journal of Psychosomatic Research, 62*, 197–205.

Lash, S. J., Eisler, R. M., & Schulman, R. S. (1990). Cardiovascular reactivity to stress in men: Effects of masculine gender role stress appraisal and masculine performance challenge. *Behavior Modification, 14*, 3–20.

Martz, D. M., Handley, K. B., & Eisler, R. M. (1995). The relationship between feminine gender role stress, body image, and eating disorders. *Psychology of Women Quarterly, 19*, 493–508.

Perugini, M. (2005). Predictive models of implicit and explicit attitudes. *British Journal of Social Psychology, 44*, 29–45.

Spence, J. T., Helmreich, R. L., & Stapp, J. (1974). The Personal Attributes Questionnaire: A measure of sex role stereotypes and masculinity-femininity. *Catalog of Selected Documents in Psychology, 4*, 43–44.

Steffens, M. C. (2004). Is the implicit association test immune to faking? *Experimental Psychology, 51*, 165–179.

van Well, S., Kolk, A. M., & Arrindell, W. A. (2005). Cross-cultural validity of the Masculine and Feminine Gender Role Stress Scales. *Journal of Personality Assessment, 84*, 271–278.

Wesseling, K. H., Jansen, J. R., Settels, J. J., & Schreuder, J. J. (1993). Computation of aortic flow from pressure in humans using a nonlinear, three-element model. *Journal of Applied Physiology, 74*, 2566–2573.