Positive stereotypes, negative outcomes: Reminders of the positive components of complementary gender stereotypes impair performance in counter-stereotypical tasks

Rotem Kahalon[1]* , Nurit Shnabel[1] and Julia C. Becker[2]

[1]Tel-Aviv University, Israel
[2]Osnabruck University, Germany

Gender stereotypes are complementary: Women are perceived to be communal but not agentic, whereas men are perceived to be agentic but not communal. The present research tested whether exposure to reminders of the positive components of these gender stereotypes can lead to stereotype threat and subsequent performance deficits on the complementary dimension. Study 1 ($N = 116$ female participants) revealed that compared to a control/no-stereotype condition, exposure to reminders of the stereotype about women's communality (but not to reminders of the stereotype about women's beauty) impaired women's math performance. In Study 2 ($N = 86$ male participants), reminders of the stereotype about men's agency (vs. a control/no-stereotype condition) impaired men's performance in a test of socio-emotional abilities. Consistent with previous research on stereotype threat, in both studies the effect was evident among participants with high domain identification. These findings extend our understanding of the potentially adverse implications of seemingly positive gender stereotypes.

While the message that negative stereotypes are anti-egalitarian and socially unacceptable is reinforced in contemporary Western society, positive stereotypes are prevalent and considered socially acceptable (Czopp, Kay, & Cheryan, 2015). The present research tested whether, despite their subjectively positive tone, positive gender stereotypes might lead to negative outcomes. In particular, we examined whether highlighting the positive stereotypes about women's communality and men's agency can lead to *stereotype threat* effects. Stereotype threat denotes group members' concern about confirming negative stereotypes regarding their ingroup's inferior ability (Steele, 1997). Stereotype threat is typically induced by exposure to reminders of these negative stereotypes (Steele, Spencer, & Aronson, 2002). Ironically, the fear of stereotype confirmation among members of stigmatized groups causes stress (Blascovich, Spencer, Quinn, & Steele, 2001) which affects precisely the cognitive systems and behavioural strategies required for optimal performance (Schmader, Johns, & Forbes, 2008) – resulting in systematic performance deficits (Steele *et al.*, 2002).

Correspondence should be addressed to Rotem Kahalon, The School of Psychological Sciences, Tel-Aviv University, Ramat Aviv, Tel-Aviv 69978, Israel (email: rotemkah@mail.tau.ac.il).

In the case of gender stereotypes, accumulating evidence shows that unless the negative stereotype about their gender is explicitly refuted (Spencer, Steele, & Quinn, 1999), taking math examinations can be a psychologically threatening experience for female students, resulting in impaired math performance (see Walton & Spencer, 2009, for a meta-analysis). Similar effects were obtained when examining women's performance in other stigmatized domains, such as driving (Yeung & von Hippel, 2008). Interestingly, even when neutral tasks in non-stigmatized domains (e.g., non-verbal reasoning) were described as producing gender differences which favoured men, women's subsequent performance on these tasks was impaired (Pavlova, Weber, Simoes, & Sokolov, 2014). These findings show that, besides the negative stereotypes about women's inferior ability in particular domains (such as science, technology, engineering, and math; Davies, Spencer, Quinn, & Gerhardstein, 2002), there is also a broader negative stereotype about women's general incompetence (Glick & Fiske, 2001), which may lead to stereotype threat effects.

With regard to stereotype threat among men, research revealed that men's performance in a social sensitivity test was impaired when participants were told that the test assessed social sensitivity – a domain in which men are stereotypically perceived to have inferior ability than women (Briton & Hall, 1995) – yet improved when participants were told that the test measured information processing (Koenig & Eagly, 2005). In a similar vein, male participants who were reminded that men are not as good as women in processing affective information made more errors in classifying affective (but not non-affective) words in a lexical decision task than did men who were not reminded of this negative stereotype (Leyens, Désert, Croizet, & Darcis, 2000). Taken together, the above studies demonstrate that negative gender stereotypes operate as self-fulfilling prophecies, undermining women's and men's performance in counter-stereotypical tasks.

According to the Integrated Process Model of Stereotype Threat (Schmader *et al.*, 2008), the detrimental consequences of stereotype threat on performance should be particularly pronounced in the three following conditions. The first is when one's *group membership becomes salient*. To illustrate, when women took a difficult math test while having a solo status (i.e., being the only woman present in the room), their performance was impaired as compared to women who took the same test in the presence of other women (Inzlicht & Ben-Zeev, 2000). The second type of circumstances is when *ego involvement is high*. This leads to the ironic result that those who care most about doing well in the stigmatized domain are more vulnerable to stereotype threat. For example, women with high math identification (who perceive math as internally and externally rewarding; Smith & White, 2001) were found to suffer the most from the effects of stereotype threat both psychologically (e.g., bifurcation of their feminine identity; Pronin, Steele, & Ross, 2004) and in terms of their math performance (Good, Aronson, & Harder, 2008; Keller, 2007). The third type of circumstances that increase susceptibility to stereotype threat is when *the stereotype's salience is high*. For example, describing a given math test as producing gender differences in favour of men substantially worsened female participants' test performance (Spencer *et al.*, 1999).

Previous research has found that making negative gender stereotypes (e.g., about women's math incompetence, Spencer *et al.*, 1999; or men's socio-emotional insensitivity, Koenig & Eagly, 2005) salient has a detrimental impact on women's and men's performance in the stereotyped domain. The goal of this study was to test the novel hypothesis that heightened salience of *positive* gender stereotypes can also lead to impaired performance among women and men in tasks in which their gender is

stereotypically perceived to have inferior ability. This hypothesis is based on previous research which has found that stereotypes about groups are often compensatory (Kervyn, Yzerbyt, Judd, & Nunes, 2009).

The complementary nature of stereotypes

Accumulating converging evidence suggests that there are two fundamental content dimensions along which social targets (i.e., individuals and groups) are perceived and judged (Abele & Wojciszke, 2013). One dimension represents traits such as good nature, morality, trustworthiness, nurturance, and sociability, whereas the other dimension represents traits such as intelligence, capableness, status, confidence, ambition, dominance, and power (Fiske, Cuddy, & Glick, 2007). Different theories use different labels to denote these dimensions. For example, Resources Theory (Foa & Foa, 1980) uses the terms 'status' and 'love', the Big Two theory (Abele & Wojciszke, 2013) uses the terms 'agency' and 'communion', and the Stereotype Content Model (Fiske, Cuddy, Glick, & Xu, 2002) uses the terms 'competence' and 'warmth' when referring to these different dimensions. Despite these differences, it has been argued that these labels describe 'what boil down to virtually the same two dimensions' (Cuddy, Fiske, & Glick, 2008, p. 65; see also Kervyn, Yzerbyt, & Judd, 2010).

Most relevant to the purposes of the present research is the Stereotype Content Model. According to this model, stereotypes fall into four quadrants reflecting different combinations of warmth and competence. For example, people often stereotypically perceive their own ingroup to be high on both these dimensions, whereas other groups, such as drug addicts and homeless people, are perceived to be low on both competence and warmth (Fiske *et al.*, 2002). The model further argues that the content of group stereotypes is often mixed, such that groups that are stereotypically perceived to be competent are also perceived to be cold and immoral, whereas groups that are stereotypically perceived to be high on warmth are also perceived to be incompetent and dependent. Hence, opposite to the judgement of other individuals, for whom the two dimensions correlate positively (resulting in a 'halo effect'; Rosenberg, Nelson, & Vivekananthan, 1968), when judging other groups 'warmth and competence often correlate negatively' (Fiske *et al.*, 2007, p. 79) – resulting in complementary stereotypes.

Gender stereotypes, which portray men as agentic and women as communal (e.g., Eagly & Wood, 1999; see Rudman & Glick, 2008, for a review), provide a striking example of such complementarity. With the exception of some subgroups (e.g., feminists, or gay men; Fiske *et al.*, 2002), men are stereotypically perceived to be dominant, ambitious, strong, competitive, independent, and good in abstract thinking and problem-solving, whereas women are stereotypically perceived to be nurturing, warm, caring, and able to provide emotional support to others due to their emotionality and understanding (Cejka & Eagly, 1999; Deaux & Lewis, 1984). The perception that men and women possess complementary traits reinforces gender inequality because it supports the belief that men are suitable to the public sphere whereas women are suitable to the private, domestic sphere (Glick & Fiske, 2001).

This understanding lies at the heart of theorizing on *benevolent sexism* that, as opposed to overtly hostile sexism, puts women on a metaphorical pedestal by highlighting their kindness, purity and morality (Glick & Fiske, 1996). Despite its seemingly positive tone, by stressing the interdependency and complementarity of the genders, the ideology of benevolent sexism subtly implies that women need men's protection and provision because they lack agency (Glick & Fiske, 2001). The

corresponding view that men are 'bad but bold' similarly functions to support the existing gender inequality, by characterizing men as being naturally designed for dominance (Glick *et al.*, 2004). In line with this theorizing, empirical evidence has found that exposure to complementary gender stereotypes increased people's justification of the existing gender system (Jost & Kay, 2005).

The present research

The present research puts forward the argument that, besides their contribution to legitimizing traditional gender roles, an additional means through which complementary gender stereotypes reinforce these roles is by undermining women's and men's performance in negatively stereotyped domains. This argument is based on findings which show that negative stereotypes can be activated through reminders of complementary positive stereotypes (Fiske *et al.*, 2015). Specifically, research on the 'innuendo effect' (Kervyn, Bergsieker, & Fiske, 2012) has found that when participants are exposed to descriptions of social targets that focus on their positive traits (e.g., describing a target as high on warmth), they interpret these descriptions as implying complementary negative information as well (e.g., that the target is also low on competence).

Based on these findings, which underline the compensatory nature of stereotypes, we hypothesized that exposure to reminders of positive gender stereotypes can lead to stereotype threat effects. Study 1 examined whether exposure to the stereotype about women's communality would impair female participants' performance in a math test. Study 2 investigated whether exposure to the stereotype about men's agency would impair male participants' performance in a test of their socio-emotional abilities.

In line with previous findings (e.g., Keller, 2007), we expected these effects to be moderated by domain identification, defined as perceiving the domain as attractive, important, feasible, and having favourable internal and external outcomes (Smith & White, 2001). People who are highly identified with a particular domain tend to have greater success in this domain (Smith & White, 2001), which constitutes a main ingredient in the global self-evaluation process (Osborne, 1997). Ironically, this is exactly what makes high identifiers susceptible to stereotype threat effects (Schmader *et al.*, 2008). We therefore expected that the effects of exposure to positive stereotypes would occur among female (Study 1) and male (Study 2) participants who highly identified with the domains of math and socio-emotional ability (respectively). Together, these studies aimed to demonstrate how positive gender stereotypes can have negative consequences for performance in counter-stereotypical tasks, and shed light on the characteristics of women and men who are most likely to suffer these consequences.

STUDY 1

The goal of Study 1 was to test whether exposure to a reminder of the positive stereotype about women's communality (i.e., warmth and nurturance) can lead to impaired math performance among female participants with high math identification. Previous research has found that exposure to manifestations of benevolent sexism (e.g., benevolently sexist comments in the context of a job interview, compared to hostile sexist or neutral comments) undermines women's sense of competence (Dumont, Sarlet, & Dardenne, 2010) and leads them to describe themselves as less task-oriented (Barreto, Ellemers, Piebinga, & Moya, 2010). Moreover, when women were treated in a patronizing manner

by their male superior – who gave them praise but few valued resources – they felt angry and performed worse in a subsequent task from a traditionally masculine domain (Vescio, Gervais, Snyder, & Hoover, 2005). In another study, receiving a patronizing comment from a male recruiter, who told the female candidate that the men with whom she was supposed to work had 'agreed to give you time and help' (p. 767), led to poorer performance on a subsequent problem-solving task, an effect mediated by the mental intrusions women experienced about their sense of competence (Dardenne, Dumont, & Bollier, 2007).

The latter two studies (Dardenne *et al.*, 2007; Vescio *et al.*, 2005) are generally consistent with our theorizing that exposure to stereotypes which portray women as nice and likeable – and hence deserving to be treated in a chivalrous manner (Glick & Fiske, 2001) – can lead to the performance deficits associated with stereotype threat. Yet, these studies examined the effects on performance of receiving patronizing treatment, not of exposure to stereotypes about women's communality. In both studies, the source of benevolent sexism was a man who was superior to the female participant, the benevolent sexist behaviour was targeted directly at the participant, and the supposedly 'kind' behaviour was patronizing, subtly implying that the participant is viewed as incompetent and non-agentic (hence she was obviously in need for help from her male colleagues, Dardenne *et al.*, 2007; or not provided with resources, Vescio *et al.*, 2005). The main goal of Study 1 was to examine whether exposure to the stereotype about women's communality, when (1) done outside of a context of gender hierarchy in which the female participant is in an inferior position, and (2) without directly relating these stereotypes to the participant herself, would be sufficient to induce women with stereotype threat and subsequent impaired math performance.

An additional, exploratory goal of Study 1 was to examine the impact on math performance of exposing women to the beauty stereotype, which highlights women's prettiness (Cejka & Eagly, 1999) and attractiveness (Carpenter & Trentman, 1998). These characteristics are positive yet clearly distinct from the communion dimension (Fletcher, Simpson, Thomas, & Giles, 1999). Hence, examining the effect of exposure to the beauty stereotype allowed us to determine whether any positive gender stereotype may lead to performance decrements, or whether such decrements are limited to stereotypes about women's warmth and communality – which have a compensatory nature (i.e., are known to be associated with incompetence, Kervyn *et al.*, 2010). Importantly, in this regard, past research has shown no consistent link between beauty and (in)competence: While there is evidence of a positive link between perceptions of beauty and competence (e.g., Jackson, Hunter, & Hodge, 1995), there is also evidence of a negative association (e.g., Heilman & Stopeck, 1985) in line with the 'beautiful but dumb' stereotype.

We used a three-cell experimental design such that following a measure of their math identification (Smith & White, 2001), female participants were either assigned to a condition which exposed them to the communality stereotype (highlighting women's nurturance qualities), the beauty stereotype (highlighting women's attractiveness), or a no-stereotype/control condition. They then took a difficult math test. With regard to the effect of the beauty stereotype condition, given the exploratory nature of our investigation, we did not have an *a priori* prediction. However, we did predict that compared to the control condition, participants in the communality-stereotype condition, especially those participants high (rather than

low) in math identification, would show impaired math performance (i.e., a two-way interaction).

Pilot study

Before running Study 1, we conducted a pilot study in order to confirm that the communality-stereotype manipulation induced stereotype threat among women. Participants in this pilot were 86 female undergraduate students of psychology (M_{age} = 23.20, SD = 2.57) who were recruited through advertisements on campus and on the Psychology School's website. Participants were Israeli Jews, and their native tongue was Hebrew; 86% described themselves as heterosexual, and the rest as either gay (2%), bisexual (4%), or did not want to report (8%). Participants were randomly assigned either to the control/no-stereotype or the communality-stereotype conditions. After reading the text that constituted the experimental manipulation and answering three reading comprehension questions (see Method below), participants were ostensibly told that they were about to take a brief math quiz and were given several example questions. Next, participants were asked to respond to three 7-point items (1 = *not at all* to 7 = *very much*) taken from Marx's (2012) threat appraisal scale, which measures participants' experience of stereotype threat (e.g., 'I worry that my ability to perform well on math tests is affected by my gender' and 'I worry that, because I know the negative stereotype about women and math, my anxiety about confirming that stereotype will negatively influence how I perform on math tests'; α = .80). After completing the threat appraisal measure, participants were thanked and debriefed.

A regression analysis with the experimental condition (dummy coded), math identification (standardized), and their two-way interaction as predictors was found to be significant, F (3, 82) = 2.98, p = .036. As intended, participants in the communality-stereotype condition (M = 2.20, SD = 1.34) showed higher levels of stereotype threat compared to participants in the control condition (M = 1.58, SD = 1.05), β = .22, t = 2.01, p = .048. Neither the effect of math identification, nor its interaction with the experimental condition reached significance, ps > .228. The pilot study confirmed that the communality-stereotype manipulation used in Study 1 successfully induced stereotype threat among women.

Method

Participants

Due to feasibility constraints, the sample size of Study 1 was set to 120 participants (40 participants per experimental cell). An *a priori* power analysis using G*Power calculator (Faul, Erdfelder, Buchner, & Lang, 2009) revealed that this sample size allowed to detect a medium effect size (f^2 = .12) at a 5% two-sided significance level and power of 80%. Participants were female students of a large Israeli university, who were recruited through advertisements on campus and on the Psychology School's website. They received either course credit or 40 NIS for participation. Four participants were excluded from analysis because they reported experiencing many interruptions during the study (excessive noise due to construction in the adjacent laboratory). The final sample thus included 116 participants.[1] All the participants were

---

[1] *The key communality (vs. control) condition × math identification interaction remained significant, β = −.27, t (113) = −2.22, p = .029, when all 120 participants were included in the analysis.*

Israeli Jews, and their native tongue was Hebrew; 91% described themselves as heterosexual, and the rest as either gay (5%) or bisexual (4%);[2] M_{age} = 23.47 (*SD* = 2.78).

Procedure and materials

Participants were invited to take part in a laboratory study on 'academic abilities in different domains'. They did not know in advance that they were about to take a math test. All the materials (i.e., instructions, manipulations, and measures) were computerized. Upon their arrival to the laboratory, participants completed a demographic questionnaire, which included a question about their psychometric score (the Israeli equivalent of the SAT score) – an available proxy of their pre-existing math ability, and a five-item measure (adapted from Smith & White, 2001) of their math identification. This measure aims to capture participants' perceived enthusiasm, interest, and success in math (e.g., 'I enjoy math and math-related fields', 'It is highly likely that I will work in a math-related field'; 1 = *strongly disagree* to 5 = *strongly agree*), α = .87. To disguise the purpose of the study, participants completed additional filler questions about their identification with the domain of verbal ability. Next, ostensibly presented as a verbal ability task, participants read short texts that constituted the experimental manipulation (see Appendix A). The texts, which discussed 'symbols and their meaning', were presented as taken from scientific articles and were followed by three reading comprehension questions.

The text in the control condition discussed the meaning of the Yin and Yang symbols in ancient Chinese philosophy, metaphysics, and medicine. For example, it explained that Yang represents 'the bright side of the hill' and is associated with morning, sun, and energy, whereas Yin represents 'the dark side of the hill' and is associated with night, moon, and freeze. The text in the beauty stereotypes condition discussed the meaning of the Mars and Venus symbols, representing male and female gender symbols (respectively). For example, it explained that the Venus symbol is shaped like a hand mirror, reflecting the importance of beauty for women, and discussed the evolutionary perspective on women's attractiveness as an indicator of health and fertility. The text in the communality-stereotype condition discussed the meaning of the circle and the square symbols in the Bender–Gestalt test, representing femininity and masculinity (respectively). For example, because communality is strongly associated with motherhood (Cuddy, Fiske, & Glick, 2004; Ridgeway & Correll, 2004), it was explained that the circle represents women's 'innate maternal qualities', such as containment and natural sensitivity. It was further explained that due to these innate characteristics, women tend to choose occupations that involve nurturing, such as nursing, social work, and working with children.

These particular experimental manipulations were chosen because they were relatively similar and comparable – all conditions had the same format such that they presented participants with a symbol consisting of two complementary elements, followed by a short scientific text that explained the meaning of this symbol and reading comprehension questions (which were included in order to verify that the participants carefully read the text). This allowed us to disguise the real purpose of the manipulation as

---

[2] *Because the traditional gender stereotypes do not apply to sexual minorities (lesbians are perceived as low on warmth and gay men are perceived as higher on warmth than on competence; Fiske* et al., *2002), we repeated the same statistical analyses without non-heterosexual participants. The critical communality (vs. control) condition × math identification interaction persisted in Study 1, β = −.32, t (103) = −2.52, p = .013, and so did the agency (vs. control) condition × domain identification interaction in Study 2, β = −.33, t (73) = 2.01, p = .049.*

'a test of verbal abilities' and activate the relevant stereotypes in a subtle manner. Subtle stereotype activation was important because stereotype threat effects are weaker when manipulated in an overt, explicit manner (Nguyen & Ryan, 2008), which can also cause reactance (e.g., Kray, Reb, Galinsky, & Thompson, 2004).

Following the assignment of participants to the three experimental conditions, they completed a manipulation check to verify that the salience of the stereotypes increased as intended. For this purpose, participants completed a shortened version of the Twenty Statements Test (TST, Bugental & Zelen, 1950; Kuhn & McPartland, 1954), a self-concept measure in which participants were asked to make ten different statements about themselves by completing the sentence 'I am. . .'. Two independent coders classified the free responses into four categories: (1) traits or roles related to the communal dimension, representing warmth, morality, and sociability (e.g., 'I am an empathetic person', 'I am a caring sister'), (2) traits related to one's agency, competence, and ambitiousness (e.g., 'I am assertive', 'I am going to be a successful person'), (3) traits related to one's physical appearance (e.g., 'I am tall', 'I am fat'), and (4) other (e.g., 'I am a waitress', 'I am bored'). Cohen's kappa was .68, representing good agreement (Landis & Koch, 1977). Cases of disagreement were resolved by a third coder. Because activating stereotypes about one's ingroup increases the accessibility of traits that are congruent with these stereotypes in one's self-concept (Bry, Follenfant, & Meyer, 2008), we expected that, compared to the control condition, participants in the communality-stereotype condition would write more sentences about their communal traits and participants in the beauty stereotype condition would write more sentences about their physical appearance.

Next, participants completed two difficult math tests. Adapted from Shnabel, Purdie-Vaughns, Cook, Garcia, and Cohen (2013), the goal of the first test was to induce participants with stereotype threat, by exposing them to ten extremely difficult questions that were nearly unsolvable.[3] The second test was similar to the GRE math examination. It lasted 15 min and included 14 difficult yet solvable questions. Participants earned one point for each correct answer. Upon completion, participants were thanked and debriefed.

The protocols and data files can be accessed either through the Open Science Framework, https://osf.io/sa9r2, or upon email request from the first author.

Results

Self-concept (TST)

As intended, the number of statements that related to participants' communal traits was significantly higher in the communality-stereotype condition ($M = 2.15$, $SD = 1.44$) compared to the beauty stereotype condition ($M = 1.34$, $SD = 1.12$), $p = .012$, or the control condition ($M = 1.41$, $SD = 1.07$), $p = .023$. No difference was found between the control and beauty stereotype conditions, $p = .968$.

Similarly, the number of statements that related to participants' physical appearance was significantly higher in the beauty stereotype condition ($M = 0.71$, $SD = 1.11$) compared to the communality-stereotype condition ($M = 0.21$, $SD = 0.57$), $p = .009$, or

---

[3] *Supporting the claim that, as intended, this test was nearly unsolvable, while participants' pre-existing math ability (i.e., psychometric score) predicted their score in the second test (see Table 1), it failed to predict their score in the first test, in which they had to guess the answers, $\beta = .09$, t $(109) = .96$, p = .341. In addition, the mean score in the first test was only 4.44, SD = 2.12, whereas the mean score in the second test was 10.68 (SD = 2.10).*

the control condition (M = 0.08, SD = 0.27), p = .001. No difference was found between the communality stereotype and control conditions, p = .722.

Table 1. Results of regression analysis on performance in the math test (Study 1)

Predictors	B	SE	β	t	p
Intercept	5.008	1.533		3.266	.001
Pre-existing math ability	.009	.002	.330	3.803	.000
Exposure to communality stereotype (vs. control)	−.302	.437	−.068	−.691	.491
Exposure to beauty stereotype (vs. control)	−.217	.440	−.049	−.493	.623
Math identification	1.133	.323	.532	3.514	.001
Exposure to communality stereotype × math identification	−1.202	.446	−.332	−2.692	.008
Exposure to beauty stereotype × math identification	−.056	.466	−.014	−.119	.905

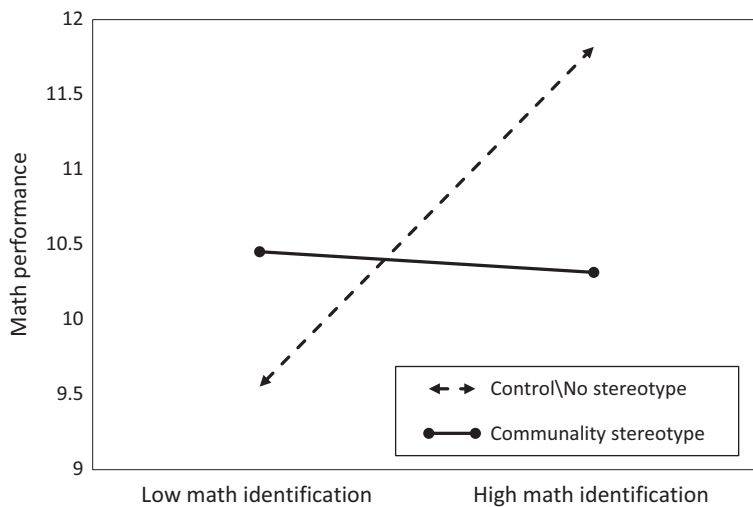*Note.* N = 116 female participants; R^2 = .28.

Figure 1. Number of correct answers in the math test among female participants with high (+1 SD) and low (−1 SD) math identification in the control/no-stereotype and communality-stereotype conditions (Study 1).

Math performance

To test our main hypothesis, we conducted a regression analysis in which math performance was the dependent variable. Besides participants' pre-existing math ability (i.e., psychometric score), the predictors were the experimental conditions (dummy coded into two contrasts such that the control condition was the reference category), math identification (standardized), and their two-way interactions. The regression model was significant, $F(6, 109) = 6.95$, $p < .001$. As seen in Table 1, consistent with previous research (e.g., Steinberg, Okun, & Aiken, 2012), pre-existing math ability and math identification predicted better performance. As expected, the communality (vs. control)

condition × math identification interaction was also significant.[4] Simple slopes analysis revealed that exposure to the communality stereotype did not affect the math performance of participants with low math identification (1 *SD* below average), $B = .90, p = .117$, but as expected, exposure to the communality stereotype impaired the performance of participants with high math identification (1 *SD* above average), $B = -1.50, p = .028$. Figure 1 illustrates the obtained pattern of results. No other effects or interactions were significant.

Discussion

Study 1's findings are fully consistent with our prediction that exposure to the positive stereotype about women's communality would induce stereotype threat among female participants, resulting in impaired math performance particularly among those who care most about doing well in this domain (i.e., women with high math identification). Interestingly, exposure to the beauty stereotype did not lead to a similar impairment in math performance. While one should be cautious with interpreting null effects, this finding is consistent with the possibility that stereotype threat is not activated in response to any positive gender stereotype, but rather specifically in response to compensatory stereotypes – in this case, the 'warm but dumb' stereotype (Fiske, 2012).

Despite the support for our theorizing, one limitation of Study 1 is that perhaps participants' exposure to extremely difficult questions in the first part of the math test induced them with stereotype threat even in the control condition. Importantly, however, stereotype threat does not take an all or none form (Schmader *et al.*, 2008) – one could argue that even though participants in the beauty stereotype and the control conditions experienced *some* threat, participants in the communality-stereotype condition experienced *greater* threat. If anything, this made it harder to demonstrate performance decrements in the communality-stereotype condition (because a certain degree of decrement has occurred even in the other two conditions). Nevertheless, we acknowledge that future research should try and replicate the results without using the extremely difficult first part of the math test, allowing a 'cleaner' manipulation.

STUDY 2

Study 2 extended Study 1 by testing the corresponding hypotheses among men – who are susceptible to stereotype threat when performing tasks that evaluate their socio-emotional abilities (Koenig & Eagly, 2005; Leyens *et al.*, 2000). For this purpose, male participants of Study 2 were first assigned to either a control/no-stereotype condition or to a condition that reminded them of the positive stereotype about men's agency (i.e., leadership and assertiveness) and then took a test that evaluated their ability to identify facial expressions of emotions. We tested the prediction that – especially among men with

---

[4] *Adding participants' score in the first math test as an additional predictor did not significantly improve the model, $\Delta R^2 = .012$, $F_{change} (7, 108) = 1.87$, p = .175, and the effect of participants' score in the first test on their score in the second test was non-significant, $\beta = .13$, t (108) = 1.37, p = .175. Also, in a regression model with the predictors reported in Table 1 yet with performance in the first (rather than the second) math test as the dependent variable, neither the effect of the communality (vs. control) condition nor its interaction with math identification reached significance, $\beta s < |.10|$, ps > .341. Finally, when participants' pre-existing math ability was not included as a predictor in the model, the key communality (vs. control) condition × identification interaction was only marginally significant, $\beta = -.22$, t (110) = -1.70, p = .091. However, using this covariate was justified by the a priori knowledge that psychometric scores (our proxy of pre-existing math ability) are strong predictor of academic achievement (Oren, Kennet-Cohen, Turvall, & Allalouf, 2014; Steinberg et al., 2012).*

high domain identification, for whom having high socio-emotional abilities is important and rewarding – exposure to a reminder of the agency stereotype would lead to impaired test performance.

Method

Participants

The sample size was set to 80 participants (40 participants per experimental cell). An *a priori* power analysis using G*Power calculator (Faul *et al.*, 2009) revealed that this sample size allowed to detect a medium effect size (f^2 = .14) at a 5% two-sided significance level and power of 80%. Due to overbooking, we ended up recruiting 91 male participants through snowball sampling. Five participants were excluded from analysis because, in response to a question at the end of the study, they indicated experiencing many interruptions during the study.[5] The final sample thus included 86 participants, M_{age} = 25.58, SD = 3.72, who volunteered to complete a web-based questionnaire. All the participants were Israeli Jews, and their native tongue was Hebrew. Of the participants, 90% described themselves as heterosexual and 10% as gay.

Procedure and materials

The procedure generally matched that of Study 1. Participants were invited to take part in a study on 'abilities in different domains'. They did not know that they were about to take a test measuring their socio-emotional abilities. All the materials were computerized. Participants conducted the experiment from their home, in a prescheduled time. A couple of minutes prior to the time in which the experiment was scheduled to begin, the experimenter called the participant to verify that he was online and ready to begin.

Participants first completed a demographic questionnaire, which included a four-item measure of their domain identification (adapted from Smith & White, 2001; to fit the domain of socio-emotional abilities; for example, 'I find myself very interested in what other people think or feel', 'It is highly likely that my future career will involve understanding the feelings and expressions of other people'; α = .71). Next, ostensibly presented as a verbal ability task, participants read the short texts that constituted the experimental manipulation (see Appendix B). The control condition was identical to that used in Study 1. The agency-stereotype condition was similar to the communality condition used in Study 1, in that it discussed the meaning of the circle and the square symbols in the Bender–Gestalt test, yet it focused on the qualities of masculinity (represented by the square). For example, it was explained that the angled shape of the square represents men's typical qualities – such as assertiveness, strength, and ambitiousness – which make them especially suitable for leadership positions.

After answering three reading comprehension questions about the text to which they were exposed, participants completed a shortened version of the TST (to verify that salience of the agency stereotype increased as intended, see Study 1). Two independent coders classified the free responses into three categories: (1) traits or roles related to one's agency and ambitiousness, (2) traits or roles related to one's warmth and communality, and (3) other. Cohen's kappa was .70 (representing good agreement; Landis & Koch, 1977). Disagreements were resolved by a third coder.

---

[5] *The key agency (vs. control) condition × domain identification interaction remained significant, β = −.36, t (87) = −2.13, p = .036, when including all 91 participants in the analysis.*

As the primary outcome variable, participants completed the Adult Facial Expressions subtest from the Diagnostic Analysis of Nonverbal Accuracy 2 (DANVA2; Nowicki & Duke, 1994), a well-established instrument designed to measure the ability to apprehend emotions of others. Understanding others' emotions is a key component of socio-emotional intelligence (Mayer, Salovey, & Caruso, 2004), and it is also the most reliably validated facet of socio-emotional ability (Elfenbein & Ambady, 2002). Specifically, the Adult Facial Expressions subtest examines emotion recognition and it consists of 24 facial photographs of an equal number of female and male adults making one of four expressions. Participants were asked to choose the correct feeling expressed in the facial expression as fast as they could. They earned one point for each correct answer.

Following the test, participants responded to two 7-point threat appraisal items (1 = *not at all* to 7 = *very much*) that evaluated their experience of stereotype threat (adapted from Marx, 2012): 'I am afraid that my ability to perform well on the emotional recognition test was influenced from me being a man', and 'I believe that the awareness to the stigma about men's poor socio-emotional ability affected my performance on the emotion recognition test'; $r = .82$, $p < .001$. We refrained from measuring threat appraisal (the manipulation check) prior to the main outcome variable (test performance) because we were concerned that the blatant wording of this measure might expose the study's real purpose and/or induce participants to experience stereotype threat even in the control condition (by increasing the salience of the negative stereotype about men's socio-emotional ability). This reversal of measures has been used in other social psychological research when the regular order is problematic (e.g., Shnabel, Bar-Anan, Kende, Bareket, & Lazar, 2016). Upon completion, participants were thanked and debriefed.

Results

Self-concept (TST)
Our manipulation check revealed that, as intended, participants generated more agency-related statements in the agency-stereotype condition ($M = 2.46$, $SD = 1.64$) than in the control condition ($M = 1.17$, $SD = 1.26$), $t(84) = 4.01$, $p < .001$.

Threat appraisal
We conducted a regression analysis with experimental condition (dummy coded), domain identification (standardized), and their two-way interaction as predictors. Although the model was not significant, $F(3, 82) = 1.39$, $p = .251$, the effect of the experimental condition was significant such that participants in the agency-stereotype condition ($M = 2.23$, $SD = 1.34$) reported experiencing more threat than participants in the control condition ($M = 1.67$, $SD = 1.13$), $\beta = .22$, $t = 2.02$, $p = .047$. Neither the effect of domain identification nor its interaction with the experimental condition reached significance, ps > .777.

Task performance
To test our main hypothesis, we conducted a regression analysis in which participants' score for the emotion recognition task was the dependent variable, and the predictors were the experimental condition (dummy coded), domain identification (standardized),

Table 2. Results of regression analysis on performance in the test of socio-emotional ability (Study 2)

Predictors	B	SE	β	t	p
Intercept	17.71	.340		52.13	.000
Exposure to agency stereotype (vs. control)	−.119	.464	−.027	−.256	.798
Domain identification	.740	.371	.321	1.993	.050
Exposure to agency stereotype × domain identification	−1.320	.492	−.433	−2.682	.009

Note. $N = 86$ male participants; $R^2 = .082$.
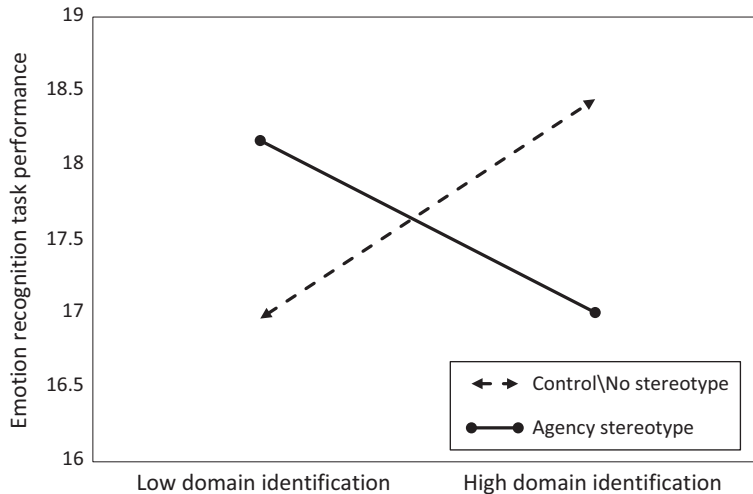
Figure 2. Number of correct answers in the emotion recognition test among male participants with high (+1 *SD*) and low (−1 *SD*) domain identification in the control/no-stereotype and agency-stereotype conditions (Study 2).

and their two-way interaction. The obtained regression model was marginally significant, $F(3,82) = 2.44$, $p = .070$. As seen in Table 2 and consistent with Study 1, domain identification was associated with better test performance. In addition, the expected two-way interaction between domain identification and the experimental condition was significant. Simple slopes analysis revealed that exposure to the agency stereotype had an unexpected positive effect on test performance among men with low domain identification (i.e., 1 *SD* below average), $B = 1.20$, $p = .084$. Given its marginal significance and the fact that we did not find a corresponding effect among low identifiers in Study 1, this effect should be treated cautiously. Importantly, and as expected, exposure to the agency stereotype significantly impaired the test performance of men with high domain identification (i.e., 1 *SD* above average), $B = −1.44$, $p = .034$, see Figure 2.

Discussion

In line with our predictions, Study 2 revealed that exposure to the positive stereotype about men's agency induced male participants with stereotype threat, resulting in impaired performance in a test of emotion recognition – an important facet of socio-emotional ability. Consistent with Study 1's results and the stereotype threat literature

(see Schmader *et al.*, 2008), performance decrements were evident specifically among men for whom doing well in this domain was highly important.

Despite the support for our predictions, a limitation of Study 2 is that the results can be alternatively explained by participants' experience of masculinity threat, rather than stereotype threat. Because manhood is a precarious (i.e., elusive and tenuous) status, men respond to threats to their masculinity (e.g., information according to which they possess feminine traits) by exhibiting hypermasculine behaviour, such as physical aggression (Vandello & Bosson, 2013). Bearing in mind that positive stereotypes are prescriptive (Glick & Rudman, 2010), it is possible that exposure to the agency stereotype induced male participants – especially those who identify with the (traditionally feminine) socio-emotional domain – with the feeling that they are not sufficiently masculine. Feeling that they should be ambitious and agentic rather than empathic and communal, in turn, could be the reason for their impaired test performance.

Nevertheless, that participants in the positive stereotype condition reported higher threat appraisals (Marx, 2012) is consistent with a stereotype threat explanation. Moreover, domain identification is a well-established moderator in the stereotype threat literature (Schmader *et al.*, 2008) but not in the precarious manhood literature. Finally, stereotype threat is a more parsimonious explanation for our findings, because it can account for the results of both Study 1 and Study 2. Still, exploring the differences between stereotype and masculinity threat effects on men may be an intriguing avenue for future research.

GENERAL DISCUSSION

Two experiments revealed that exposure to reminders of the positive components of complementary gender stereotypes can lead to stereotype threat and consequent performance deficits. In particular, Study 1 revealed that exposure to the stereotype about women's communality (but not about women's beauty) impaired the math performance of female participants who strongly identified with math. Study 2 found that exposure to the stereotype about men's agency impaired performance in a test of socio-emotional ability among male participants who strongly identified with the socio-emotional domain. The finding that women and men who cared about doing well in these counter-stereotypical tasks suffered from greater performance decrements is consistent with the stereotype threat literature: Individuals with high domain identification are especially motivated to disconfirm the negative stereotype about their group, yet ironically this causes extra counterproductive stress (Schmader *et al.*, 2008).

Previous research on gender-based stereotype threat has highlighted the detrimental consequences for performance of activating *negative* stereotypes about women's inferior STEM abilities (e.g., Spencer *et al.*, 1999) and men's inferior socio-emotional ability (e.g., Koenig & Eagly, 2005). The present research extends this previous work by demonstrating that, due to the negative link between the Big Two; namely, warmth and communion on the one hand, and agency and competence on the other hand (Fiske *et al.*, 2007; Jost & Kay, 2005), activating *positive* stereotypes about women's communion and men's agency can also lead to impaired performance. To the best of our knowledge, this is the first research to demonstrate that exposure to the positive components of complementary gender stereotypes can induce participants with stereotype threat. These findings offer a novel integration between the stereotype threat literature (see Steele *et al.*, 2002) and research on the compensatory nature of stereotypes (Fiske *et al.*, 2015).

Our findings contribute to the growing understanding that positive stereotypes play a critical role in preserving group-based inequality in contemporary society (Czopp *et al.*, 2015). Admittedly, individuals who belong to groups that are the target of these positive stereotypes benefit psychologically from their endorsement, which helps them to preserve a positive self-concept despite the negative stereotypes about their ingroup (Biernat, Vescio, & Green, 1996). For example, women's endorsement of positive stereotypes about their gender is associated with better subjective well-being (Connelly & Heesacker, 2012; Hammond & Sibley, 2011).

Nonetheless, positive stereotypes can be harmful. First, they encourage the judgement of people based on their group affiliation, which can be an aversive experience for targets who wish to be seen as individuals separate from their groups (Siy & Cheryan, 2013). Second, positive stereotypes reinforce and legitimize the existing social order in a subtle, seemingly benevolent way that disarms resistance to inequality (Jost & Kay, 2005). For example, Becker and Wright (2011) found that priming benevolent sexism – which highlights women's stereotypical positive traits (nurturance and moral sensibility) – diminished women's willingness to take collective action to promote change towards gender equality; exposure to overtly hostile sexism did not lead to similar results. Finally, whereas negative gender stereotypes are descriptive, positive gender stereotypes are prescriptive (Glick & Rudman, 2010). Thus, they create an expectancy context that is likely to encourage stereotype-consistent behaviours – as stereotype-inconsistent behaviours that violate these prescriptions, such as women's assertion of dominance or men's expression of vulnerability, are socially sanctioned (Rudman, Moss-Racusin, Glick, & Phelan, 2012).

Our findings show that besides the justification of the existing gender system and prescription of stereotype-consistent behaviours, positive gender stereotypes reinforce conventional gender roles through directly undermining the performance of women and men in negatively stereotyped domains. Because individuals' occupational choices are heavily influenced by their self-schema (i.e., their perception of 'what I am and what am I capable of'; Eccles, 1987), such effects might have far-reaching consequences in terms of their influence on women's and men's career choices and paths. These negative consequences are exacerbated since those who mostly suffer from them are exactly the individuals who wish to succeed in negatively stereotyped domains, namely women who consider a math-related career and men who consider a career requiring high socio-emotional skills.

While most of the research on the negative consequences of gender-based stereotype threat has focused on women, genuine gender equality requires removing not only the obstacles preventing women's entrance to traditionally masculine domains, but also the ones preventing men's entrance to traditionally feminine domains (Budig, 2008). An advantage of the present research is that we examined both sides of this coin.

Limitations and future directions

Despite the general support for our theorizing, we acknowledge several limitations of the present research. One limitation is that we did not demonstrate that (1) men's math performance remains intact when reminded of the communality stereotype (Study 1) and (2) women's socio-emotional performance remains intact when reminded of the agency stereotype (Study 2). This renders an alternative explanation for the results. Perhaps, reminders of the concept of communion (in Study 1) primed participants with communal goals (e.g., 'be nice to others!'), reducing their pursuit of agentic goals (e.g., 'do well in math!'). Similarly, perhaps reminders of the concept of agency (in Study 2) primed participants with agentic goals (e.g., 'convince others to follow your leadership!'),

reducing their pursuit of communal goals (e.g., 'try to understand others' feelings!'). This priming process, rather than the experience of stereotype threat, might be responsible for the observed performance decrements. If so, men (too) would have shown impaired math performance in Study 1, and women (too) would have shown impaired socio-emotional performance in Study 2.

However, domain identification is a well-established moderator of stereotype threat effects (Schmader *et al.*, 2008), whereas priming effects are moderated by other variables (e.g., level of self-monitoring; DeMarree, Wheeler, & Petty, 2005). Hence, that domain identification moderated the effect of exposure to positive gender stereotypes in both studies bolsters our confidence that the observed performance decrements were caused by participants' experience of stereotype threat and are hence unique to women in Study 1 and to men in Study 2. Still, we acknowledge that future research should directly test the hypotheses that exposure to the warmth and competence stereotypes neither changes participants' goals (i.e., de-motivates them to do well), nor does it impair men's math performance and women's socio-emotional performance.

The findings for participants' threat appraisal, such that participants threat appraisal was higher among women in the communality-stereotype condition (see Pilot study) and men in agency-stereotype condition (see Study 2) as compared to the control/no-stereotype conditions, were also consistent with a stereotype threat explanation. A limitation of these findings, however, is that it is unclear why threat appraisals were not particularly pronounced among women and men with high domain identification – thus exactly corresponding to the observed performance decrements. To date, only one study has reported using the threat appraisal measurement (other studies using this measure have been retracted). Yet this study, which found that Black participants who completed a verbal test administered by a White (vs. Black) experimenter reported a higher level of threat (Marx & Goff, 2005), did not examine domain identification as a moderator. It is possible that the negative thoughts and emotions related to the experience of stereotype threat translate into more harmful consequences for high identifiers (e.g., because they invest more mental resources in thrusting these threatening thoughts and feelings). Future research is thus needed to examine the process through which the experience of stereotype threat translates into performance decrements.

Another limitation of the present research concerns the use of the Bender–Gestalt test's circle-and-square symbol to manipulate positive stereotypes. We used this particular manipulation for several reasons. First, we believed that a text about 'symbols and their meaning' would conceal the studies' real purpose better than a text that directly and explicitly discusses the communality (Study 1) and agency (Study 2) stereotypes. As mentioned earlier, concealing the studies' purpose was important because stereotype threat effects occur primarily when manipulated in a subtle, rather than an overt, manner (Nguyen & Ryan, 2008). Second, similar to the circle-and-square symbol, the Yin Yang symbol (used in the control condition) also symbolizes the notion of complementarity. Thus, participants in *all* conditions were exposed to the notion of complementarity. Nevertheless, we acknowledge that beyond complementarity *per se*, the use of the circle-and-square symbol exposed participants in the positive stereotype conditions to the more particular notion of *gender* complementarity. Future research should try to disentangle the two components, namely positive gender stereotypes versus complementary of gender roles.

Despite these limitations, our findings have theoretical and social importance. One of the challenges to the efforts to reach gender equality is that the mechanisms that impede it are often subtle (Swim & Hyers, 2009) and therefore harder to pinpoint as compared with more overt mechanisms (Barreto & Ellemers, 2005). While the notion that negative

stereotypes about women's incompetence and lack of agency can have detrimental consequences is relatively straightforward, the notion that positive stereotypes about women's communality might be similarly harmful is less intuitive. Raising awareness of the negative consequences of positive stereotypes may be critical for remedying gender achievement gaps, allowing women and men broader option choices that are not affected by gender stereotypes, to the benefit of society as a whole.

Acknowledgements

Funding statement

References

Abele, A. E., & Wojciszke, B. (2013). The big two in social judgment and behavior. *Social Psychology*, *44*, 61–62. https://doi.org/10.1027/1864-9335/a000137

Barreto, M., & Ellemers, N. (2005). The burden of benevolent sexism: How it contributes to the maintenance of gender inequalities. *European Journal of Social Psychology*, *35*, 633–642. https://doi.org/10.1002/ejsp.270

Barreto, M., Ellemers, N., Piebinga, L., & Moya, M. (2010). How nice of us and how dumb of me: The effect of exposure to benevolent sexism on women's task and relational self-descriptions. *Sex Roles*, *62*, 532–544. https://doi.org/10.1007/s11199-009-9699-0

Becker, J. C., & Wright, S. C. (2011). Yet another dark side of chivalry: Benevolent sexism undermines and hostile sexism motivates collective action for social change. *Journal of Personality and Social Psychology*, *101*, 62–77. https://doi.org/10.1037/a0022615

Biernat, M., Vescio, T. K., & Green, M. L. (1996). Selective self-stereotyping. *Journal of Personality and Social Psychology*, *71*, 1194–1209. https://doi.org/10.1037/0022-3514.71.6.1194

Blascovich, J., Spencer, S. J., Quinn, D. M., & Steele, C. M. (2001). Stereotype threat and the cardiovascular reactivity of African-Americans. *Psychological Science*, *12*, 225–229. https://doi.org/10.1111/1467-9280.00340

Briton, N. J., & Hall, J. A. (1995). Beliefs about female and male nonverbal communication. *Sex Roles*, *32*, 79–90. https://doi.org/10.1007/BF01544758

Bry, C., Follenfant, A., & Meyer, T. (2008). Blonde like me: When self-construals moderate stereotype priming effects on intellectual performance. *Journal of Experimental Social Psychology*, *44*, 751–757. https://doi.org/10.1016/j.jesp.2007.06.005

Budig, M. (2008). Feminism and the family. In J. Scott, J. Treas, & M. Richards (Eds.), *The Blackwell companion to the sociology of families* (pp. 416–434). New York, NY: John Wiley & Sons.

Bugental, J. F., & Zelen, S. L. (1950). Investigations into the 'self-concept'; the W-A-Y technique. *Journal of Personality*, *18*, 483–498. https://doi.org/10.1111/j.1467-6494.1950.tb01264.x

Carpenter, S., & Trentman, S. (1998). Subtypes of women and men: A new taxonomy and an exploratory analysis. *Journal of Social Behavior and Personality*, *13*, 679–696.

Cejka, M. A., & Eagly, A. H. (1999). Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Personality and Social Psychology Bulletin*, *25*, 413–423. https://doi.org/10.1177/0146167299025004002

Connelly, K., & Heesacker, M. (2012). Why is benevolent sexism appealing? Associations with system justification and life satisfaction. *Psychology of Women Quarterly*, *36*, 432–443. https://doi.org/10.1177/0361684312456369

Cuddy, A. J., Fiske, S. T., & Glick, P. (2004). When professionals become mothers, warmth doesn't cut the ice. *Journal of Social Issues*, *60*, 701–718. https://doi.org/10.1111/j.0022-4537.2004.00381.x

Cuddy, A. J., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in Experimental Social Psychology*, *40*, 61–149. https://doi.org/10.1016/S0065-2601(07)00002-0

Czopp, A. M., Kay, A. C., & Cheryan, S. (2015). Positive stereotypes are pervasive and powerful. *Perspectives on Psychological Science*, *10*, 451–463. https://doi.org/10.1177/1745691615588091

Dardenne, B., Dumont, M., & Bollier, T. (2007). Insidious dangers of benevolent sexism: Consequences for women's performance. *Journal of Personality and Social Psychology*, *93*, 764–779. https://doi.org/10.1037/0022-3514.93.5.764

Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin*, *28*, 1615–1628. https://doi.org/10.1177/014616702237644

Deaux, K., & Lewis, L. L. (1984). Structure of gender stereotypes: Interrelationships among components and gender label. *Journal of Personality and Social Psychology*, *46*, 991–1004. https://doi.org/10.1037/0022-3514.46.5.991

DeMarree, K. G., Wheeler, S. C., & Petty, R. E. (2005). Priming a new identity: Self-monitoring moderates the effects of nonself primes on self-judgments and behavior. *Journal of Personality and Social Psychology*, *89*, 657–671. https://doi.org/10.1037/0022-3514.89.5.657

Dumont, M., Sarlet, M., & Dardenne, B. (2010). Be too kind to a woman, she'll feel incompetent: Benevolent sexism shifts self-construal and autobiographical memories toward incompetence. *Sex Roles*, *62*, 545–553. https://doi.org/10.1007/s11199-008-9582-4

Eagly, A. H., & Wood, W. (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American Psychologist*, *54*, 408–423. https://doi.org/10.1037/0003-066X.54.6.408

Eccles, J. S. (1987). Gender roles and women's achievement related decisions. *Psychology of Women Quarterly*, *11*, 135–172. https://doi.org/10.1111/j.1471-6402.1987.tb00781.x

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*, 203–249. https://doi.org/10.1037/0033-2909.128.2.203

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fiske, S. T. (2012). Managing ambivalent prejudices: Smart-but-cold and warm-but-dumb stereotypes. *The Annals of the American Academy of Political and Social Science*, *639*, 33–48. https://doi.org/10.1177/0002716211418444

Fiske, S. T., Bergsieker, H., Constantine, V., Dupree, C., Holoien, D. S., Kervyn, N., . . . Swencionis, J. (2015). Talking up and talking down: The power of positive speaking. *Journal of Social Issues*, *71*, 834–846. https://doi.org/10.1111/josi.12152

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*, 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*, 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Fletcher, G. J., Simpson, J. A., Thomas, G., & Giles, L. (1999). Ideals in intimate relationships. *Journal of Personality and Social Psychology*, *76*, 72–89. https://doi.org/10.1037/0022-3514.76.1.72

Foa, E. B., & Foa, U. G. (1980). Resource theory: Interpersonal behavior as exchange. In K. J. Gergen, M. S. Greenberg, & R. H. Willis (Eds.), *Social exchange: Advances in theory and research* (pp. 77–94). New York, NY: Plenum Press. https://doi.org/10.1007/978-1-4613-3087-5

Glick, P., & Fiske, S. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, *70*, 491–512. https://doi.org/10.1037/0022-3514.70.3.491

Glick, P., & Fiske, S. T. (2001). An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist*, *56*, 109–118. https://doi.org/10.1037/0003-066X.56.2.109

Glick, P., Lameiras, M., Fiske, S. T., Eckes, T., Masser, B., Volpato, C., . . . Castro, Y. R. (2004). Bad but bold: Ambivalent attitudes toward men predict gender inequality in 16 nations. *Journal of Personality and Social Psychology*, *86*, 713–728. https://doi.org/10.1037/0022-3514.86.5.713

Glick, P., & Rudman, L. A. (2010). Sexism. In J. F. Dovidio, M. Hewstone, P. Glick, & V. M. Esses (Eds.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 328–344). Newbury Park, CA: Sage. https://doi.org/10.4135/9781446200919

Good, C., Aronson, J., & Harder, J. A. (2008). Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses. *Journal of Applied Developmental Psychology*, *29*, 17–28. https://doi.org/10.1016/j.appdev.2007.10.004

Hammond, M. D., & Sibley, C. G. (2011). Why are benevolent sexists happier? *Sex Roles*, *65*, 332–343. https://doi.org/10.1007/s11199-011-0017-2

Heilman, M. E., & Stopeck, M. H. (1985). Being attractive, advantage or disadvantage? Performance based evaluations and recommended personnel actions as a function of appearance, sex and job type. *Organizational Behavior and Human Decision Processes*, *35*, 202–212. https://doi.org/10.1016/0749-5978(85)90035-4

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, *11*, 365–371. https://doi.org/10.1111/1467-9280.00272

Jackson, L. A., Hunter, J. E., & Hodge, C. N. (1995). Physical attractiveness and intellectual competence: A meta-analytic review. *Social Psychology Quarterly*, *58*, 108–122. https://doi.org/10.2307/2787149

Jost, J. T., & Kay, A. C. (2005). Exposure to benevolent sexism and complementary gender stereotypes: Consequences for specific and diffuse forms of system justification. *Journal of Personality and Social Psychology*, *88*, 498–509. https://doi.org/10.1037/0022-3514.88.3.498

Keller, J. (2007). Stereotype threat in classroom settings: The interactive effect of domain identification, task difficulty and stereotype threat on female students' math's performance. *British Journal of Educational Psychology*, *77*, 323–338. https://doi.org/10.1348/000709906X113662

Kervyn, N., Bergsieker, H. B., & Fiske, S. T. (2012). The innuendo effect: Hearing the positive but inferring the negative. *Journal of Experimental Social Psychology*, *48*, 77–85. https://doi.org/10.1016/j.jesp.2011.08.001

Kervyn, N., Yzerbyt, V., & Judd, C. M. (2010). Compensation between warmth and competence: Antecedents and consequences of a negative relation between the two fundamental dimensions of social perception. *European Review of Social Psychology*, *21*, 155–187. https://doi.org/10.1080/13546805.2010.517997

Kervyn, N., Yzerbyt, V. Y., Judd, C. M., & Nunes, A. (2009). A question of compensation: The social life of the fundamental dimensions of social perception. *Journal of Personality and Social Psychology*, *96*, 828–842. https://doi.org/10.1037/a0013320

Koenig, A. M., & Eagly, A. H. (2005). Stereotype threat in men on a test of social sensitivity. *Sex Roles*, *52*, 489–496. https://doi.org/10.1007/s11199-005-3714-x

Kray, L. J., Reb, J., Galinsky, A. D., & Thompson, L. (2004). Stereotype reactance at the bargaining table: The effect of stereotype activation and power on claiming and creating value. *Personality and Social Psychology Bulletin*, *30*, 399–411. https://doi.org/10.1177/0146167203261884

Kuhn, M. H., & McPartland, T. S. (1954). An empirical investigation of self-attitudes. *American Sociological Review*, *19*, 68–76. https://doi.org/10.2307/2088175

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174. https://doi.org/10.2307/2529310

Leyens, J. P., Désert, M., Croizet, J. C., & Darcis, C. (2000). Stereotype threat: Are lower status and history of stigmatization preconditions of stereotype threat? *Personality and Social Psychology Bulletin*, *26*, 1189–1199. https://doi.org/10.1177/0146167200262002

Marx, D. M. (2012). Differentiating theories: A comparison of stereotype threat and stereotype priming effects. In M. Inzlicht & T. Schmader (Eds.), *Stereotype threat: Theory, process, and application* (pp. 124–140). New York, NY: Oxford University Press.

Marx, D. M., & Goff, P. A. (2005). Clearing the air: The effect of experimenter race on target's test performance and subjective experience. *British Journal of Social Psychology*, *44*, 645–657. https://doi.org/10.1348/014466604X17948

Mayer, J. D., Salovey, P., & Caruso, D. R. (2004). Emotional intelligence: Theory, findings, and implications. *Psychological Inquiry*, *15*, 197–215. https://doi.org/10.1207/s15327965pli1503_02

Nguyen, H. H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, *93*, 1314–1334. https://doi.org/10.1037/a0012702

Nowicki, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale. *Journal of Nonverbal Behavior*, *18*, 9–35. https://doi.org/10.1007/BF02169077

Oren, C., Kennet-Cohen, T., Turvall, E., & Allalouf, A. (2014). Demonstrating the validity of three general scores of PET in predicting higher education achievement in Israel. *Psicothema*, *26*, 117–126.

Osborne, J. W. (1997). Race and academic disidentification. *Journal of Educational Psychology*, *89*, 728–735. https://doi.org/10.1037/0022-0663.89.4.728

Pavlova, M. A., Weber, S., Simoes, E., & Sokolov, A. N. (2014). Gender stereotype susceptibility. *PLoS One*, *9*, e114802. https://doi.org/10.1371/journal.pone.0114802

Pronin, E., Steele, C. M., & Ross, L. (2004). Identity bifurcation in response to stereotype threat: Women and mathematics. *Journal of Experimental Social Psychology*, *40*, 152–168. https://doi.org/10.1016/S0022-1031(03)00088-X

Ridgeway, C. L., & Correll, S. J. (2004). Motherhood as a status characteristic. *Journal of Social Issues*, *60*, 683–700. https://doi.org/10.1111/j.0022-4537.2004.00380.x

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, *9*, 283–294. https://doi.org/10.1037/h0026086

Rudman, L. A., & Glick, P. (2008). *The social psychology of gender*. New York, NY: Guilford.

Rudman, L. A., Moss-Racusin, C. A., Glick, P., & Phelan, J. E. (2012). Reactions to vanguards: Advances in backlash theory. In P. Devine & A. Plant (Eds.), *Advances in experimental social psychology, 45* (pp. 167–227). Burlington, MA: Academic Press.

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, *115*, 336–356. https://doi.org/10.1037/0033-295X.115.2.336

Shnabel, N., Bar-Anan, Y., Kende, A., Bareket, O., & Lazar, Y. (2016). Help to perpetuate traditional gender roles: Benevolent sexism increases engagement in dependency-oriented cross-gender helping. *Journal of Personality and Social Psychology*, *110*, 55–75. https://doi.org/10.1037/pspi0000037

Shnabel, N., Purdie-Vaughns, V., Cook, J. E., Garcia, J., & Cohen, G. L. (2013). Demystifying values-affirmation interventions: Writing about social belonging is a key to buffering against identity threat. *Personality and Social Psychology Bulletin*, *39*, 663–676. https://doi.org/10.1177/0146167213480816

Siy, J. O., & Cheryan, S. (2013). When compliments fail to flatter: American individualism and responses to positive stereotypes. *Journal of Personality and Social Psychology*, *104*, 87–102. https://doi.org/10.1037/a0030183

Smith, J. L., & White, P. H. (2001). Development of the domain identification measure: A tool for investigating stereotype threat effects. *Educational and Psychological Measurement*, *61*, 1040–1057. https://doi.org/10.1177/00131640121971635

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, *35*, 4–28. https://doi.org/10.1006/jesp.1998.1373

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613–629. https://doi.org/10.1037/0003-066X.52.6.613

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. *Advances in Experimental Social Psychology*, *34*, 379–440. https://doi.org/10.1016/S0065-2601(02)80009-0

Steinberg, J. R., Okun, M. A., & Aiken, L. S. (2012). Calculus GPA and math identification as moderators of stereotype threat in highly persistent women. *Basic and Applied Social Psychology*, *34*, 534–543. https://doi.org/10.1080/01973533.2012.727319

Swim, J. K., & Hyers, L. L. (2009). Sexism. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping and discrimination* (pp. 407–430). New York, NY: Psychology Press.

Vandello, J. A., & Bosson, J. K. (2013). Hard won and easily lost: A review and synthesis of theory and research on precarious manhood. *Psychology of Men & Masculinity*, *14*, 101–113. https://doi.org/10.1037/a0029826

Vescio, T. K., Gervais, S. J., Snyder, M., & Hoover, A. (2005). Power and the creation of patronizing environments: The stereotype-based behaviors of the powerful and their effects on female performance in masculine domains. *Journal of Personality and Social Psychology*, *88*, 658–672. https://doi.org/10.1037/0022-3514.88.4.658

Walton, G. M., & Spencer, S. J. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological Science*, *20*, 1132–1139. https://doi.org/10.1111/j.1467-9280.2009.02417.x

Yeung, N. C. J., & von Hippel, C. (2008). Stereotype threat increases the likelihood that female drivers in a simulator run over jaywalkers. *Accident Analysis and Prevention*, *40*, 667–674. https://doi.org/10.1016/j.aap.2007.09.003
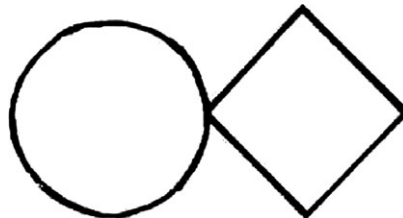
Appendix A

*The study was administered in Hebrew.
Hebrew version is available upon request from the first author.

Part A: Reading comprehension
Condition 1
Signs and their meaning

The Bender–Gestalt test is a psychological test which is performed as a part of a psychological tests battery. In this test, participants are asked to copy different figures. The first figure is comprised of a circle and a square (as seen in the picture).

The circle is the round feminine figure, while the square is the masculine figure. The circle represents the mother's ability to contain and address the child's emotional needs.

As such, it represents women's warmth and nurturance. The representation of femininity as a circle is based on the classical psychological literature, especially the renowned psychoanalyst Donald Winnicott who has coined the term 'a good enough mother' (Winnicott, 1962). Based on his clinical experience in treating hundreds of mothers and babies, Winnicott has formulated a theory which emphasizes the initial relationship between mother and child.

Winnicott has argued that a baby should not be viewed as an individual. Rather, a baby should be seen as part of a dyadic relationship with his mother. Winnicott believed that women have high emotional abilities and a motherly intuition. Specifically, he described three emotional functions of mothers:

Mirroring – reflecting to the child who he is and what he is experiencing;
Holding – holding the child physically and emotionally;
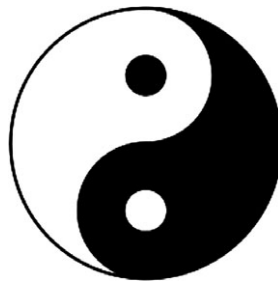Handling – taking care of the child in the face of circumstances dictated by reality.

The mother's ability to address the child's needs and contain his difficulties makes her, according to Winnicott, a 'good enough mother'. Interestingly, current research suggests that these unique feminine abilities are manifested in women's personality traits. Specifically, it was found that women are warm, sensitive and have high moral standards. In most cases, women's ability to contain the other is higher than men's. Therefore, women tend to choose occupations that involve taking care of others, such as nursing, social work, and working with children.

Please answer the following questions based on the text above:
1. Why was the circle figure chosen to represent femininity?
2. What is 'a good enough mother' according to Winnicott?
3. What are the unique personality traits which allow women to become good enough mothers?

Condition 2

Signs and their meaning

The source of the concept Yin-Yang lies in the old Chinese philosophy and metaphysics. It describes the two opposite but complementary powers which can be found in nature and among humans. The well-known symbol for Yin and Yang is called Taijitu. The mostly white part of the symbol is the Yang, and the mostly black side is the Yin. Each part contains a seed of the other – the black part in the Yang represents what is perceived to exist in a materialistic form, and the white seed in the middle of the black Yin

represents what is perceived to exist in a non-materialistic form, like 'beauty' or 'music' (Chi & Wain, 1998).

Yang means 'the lightened side of the mountain', and is connected to morning, day and live energetic actions. The Yin on the other hand means 'the shaded side of the mountain' and is connected to night and stagnation. These concepts are also linked to time. The sunset represents a time in which Yang becomes Yin. The moon and the middle of the night are full Yin, and the sunrise represents a state in which Yin becomes Yang. The noon sun represents a full Yang.

Yin and Yang are opposites; nevertheless, they support each other and are dependent on one another, they can change each other, and contain parts of each other. Unlike the dualism of bad and good, the Yin and Yang are equally important.

The Yin and Yang also serve to express an illness. Many Asian cultures manage diseases characterized by hotness/coldness, dryness/wetness with the opposite treatment. For example, diseases with symptoms characterized with extra Yang, such as nervousness and restlessness, are treated with Yin, for instance eating cold food, and vice versa.

Please answer the following questions based on the text above:
1. How are the Yin and Yang symbols represented in the time continuum?
2. How does Asian cultures treat illness according to the Yin and Yang?
3. One of the characteristics of Yin and Yang is that they contradict each other. Write two more characteristics.

Condition 3

Signs and their meaning
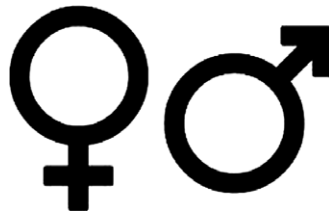
There are two well-known gender signs: The Venus sign (the love goddess sign) represent femininity – it is formed by a circle with a little cross underneath, which represents a woman's hand mirror. The lower part is the mirror handle, and the upper part (the circle) is the mirror itself. The Mars symbol, on the other hand, is constructed from a circle with an arrow coming out of it, symbolizing the god of war Mars, and therefore representing manhood.

Evolutionary studies suggest that the feminine symbols represent the importance of beauty for the female sex. Specifically, female beauty is important for survival because it is strongly connected to health and fertility. Indeed, previous research has found that men rely heavily on the visual impression of females in their first encounter with women. Thus, if a man is attracted to a woman, this usually happens immediately and quickly and is based on the physical parameters of her appearance.

Another example can be found in men's attraction to breasts, buttocks, and round thighs. The hormone responsible for fat concentration in those areas is called oestrogen. Other features indicative of health is face and body symmetry, smooth skin, and healthy

hair. Previous research has found that women with those characteristics are judged to be more attractive.

Finally, behavioural characteristics were also found to have a great impact on women's attractiveness. Examples of these characteristics include a smile (men were found to be more attracted to women with happy expressions) and dressing preferences (men are more attracted to women who dress in a way which compliments their body shape and size, for example emphasizing small waist-hip proportion). Wearing makeup and using subtle perfume were also found to affect men's attraction to women (Bordy & Mithcell, 2011).

Please answer the following questions based on the text above:

1. Why was the venus symbol chosen to symbolize women?
2. According to Brody & Mitchell, referring to women as the fairer sex has an evolutionary advantage. Give 2 examples that support this notion.
3. According to the cited studies, women can choose to practice certain behaviours in order to be perceived more attractive in men's eyes. Give two examples from the text that supports this notion.
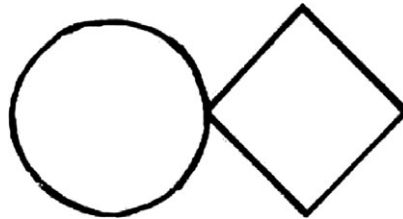
Appendix B

*The study was conducted in Hebrew and translated into English. Hebrew version is available upon request from the first author.

Part A: Reading comprehension
Condition 1
Signs and their meaning

The Bender–Gestalt test is a psychological test which is performed as a part of a psychological tests battery. In this test, participants are asked to copy different figures. The first figure is comprised of a circle and a square (as seen in the picture).

The square is the masculine figure, while the circle is the round feminine figure.

The angular figure was chosen to represent males because angularity symbolizes elements which are related to assertiveness, dominance, power, and decisiveness. It represents a person who is committed to achieving his goals. Indeed, research has demonstrated that these characteristics increase the suitability of men for management positions and their ability to negotiate successfully and effectively.

The representation of men as a square is based on the previous psychological literature which described men's characteristics in different cultures. One of the most prominent studies is that of Zeltzman (1974), who showed that manhood is not just a personal

identity but is also anchored in social arrangements. Men are perceived to be responsible for the social and familial security. A man's role is helping and protecting the mother, so that the child–mother relationship can develop without interruptions.
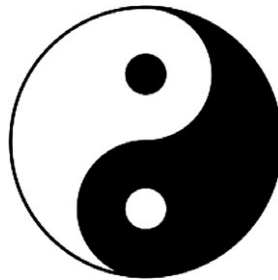
Importantly, Zeltzman's classical experiments received support from current studies examining the effects of genetics on the development of muscularity. Those studies suggest that many of the masculine characteristics described by Zeltzman have in fact a genetic origin. More specifically, these genes lead to a series of physiological changes (hormonal and others) among men which are responsible for a rise in the probability of characteristics such as dominance and analytical thinking.

Please answer the following questions based on the text above:
1. Why was the circle chosen to symbolize men?
2. What is the main finding in Zeltzman's study?
3. Does recent research support or contradict his arguments? In what way?

Condition 2

Signs and their meaning

The source of the concept Yin-Yang lies in the old Chinese philosophy and metaphysics. It describes the two opposite but complementary powers which can be found in nature and among humans. The well-known symbol for Yin and Yang is called Taijitu. The mostly white part of the symbol is the Yang, and the mostly black side is the Yin. Each part contains a seed of the other – the black part in the Yang represents what is perceived to exist in a materialistic form, and the white seed in the middle of the black Yin represents what is perceived to exist in a non-materialistic form, like 'beauty' or 'music' (Chi & Wain, 1998).

Yang means 'the lightened side of the mountain' and is connected to morning, day, and live energetic actions. The Yin on the other hand means 'the shaded side of the mountain' and is connected to night and stagnation. These concepts are also linked to time. The sunset represents a time in which Yang becomes Yin. The moon and the middle of the night are full Yin, and the sunrise represents a state in which Yin becomes Yang. The noon sun represents a full Yang.

Yin and Yang are opposites; nevertheless, they support each other and are dependent on one another, they can change each other, and contain parts of each other. Unlike the dualism of bad and good, the Yin and Yang are equally important.

The Yin and Yang also serve to express an illness. Many Asian cultures manage diseases characterized by hotness/coldness, dryness/wetness with the opposite treatment. For

example, diseases with symptoms characterized with extra Yang, such as nervousness and restlessness, are treated with Yin, for instance eating cold food, and vice versa.

Please answer the following questions based on the text above:

1. How are the Yin and Yang symbols represented in the time continuum?
2. How does Asian cultures treat illness according to the Yin and Yang?
3. One of the characteristics of Yin and Yang is that they contradict each other. Write two more characteristics.