Gender differences in self-esteem, unvarnished self-evaluation, future orientation, self-enhancement and self-derogation in a U.S. national sample

William Magee[a,*], Laura Upenieks[b]

[a] Department of Sociology, University of Toronto, Canada
[b] Department of Sociology, University of Texas at San Antonio, United States of America

A B S T R A C T

American men tend to score slightly higher on measures of self-esteem than American women. We ask whether this is because of gender differences in responsiveness to the positive and negative phrasing of self-related survey statements used to assess self-esteem. We argue that self-enhancing and self-derogatory tendencies can be inferred from wording valence effects that are common to both self-esteem and optimism. Including latent factors for those response tendencies in a bifactor measurement model transforms the latent factors for self-esteem and optimism into "unvarnished" forms of self-evaluation and future orientation. The bifactor model is shown to fit data from the National Survey of Midlife Development in the United States (MIDUS) better than a conventional measurement model. Although we observe a gender difference in self-esteem, as identified in the conventional model, no gender difference is observed in unvarnished self-evaluation identified in the bifactor model. Our results are consistent with the idea that self-esteem differs by gender due to a greater tendency for men to agree with positively worded self-statements, and a greater tendency for women to agree with negatively worded self-statements. We argue that those tendencies can be interpreted respectively as reflecting unconscious dispositions to self-enhance and self-derogate.

1. Introduction

Many studies have observed a gender difference in self-esteem (Feingold, 1994; Kling, Hyde, Showers, & Buswell, 1999; Major, Barr, Zubek, & Babey, 1999; Orth, Trzesniewski, & Robins, 2010; Rentzsch, Wenzler, & Schütz, 2016; Robins, Trzesniewski, Tracy, Gosling, & Potter, 2002), with males typically reporting slightly higher levels of self-esteem than females. Some research has investigated this gender difference by decomposing self-esteem into multiple dimensions or domains. For example, Rentzsch et al. (2016) studied gender differences in esteem related academic performance and physical appearance. Other studies have investigated whether gender differences in the effects of positively- and negatively-valenced wording of survey items contribute to the gender difference in global self-esteem scores (Michaelides et al., 2016; Salerno, Ingoglia, & Coco, 2017).[1] We build on the latter approach by decomposing self-esteem and optimism scores

into sub-dimensions comprised of positive and negative wording valence effects, as well as a sub-dimension that is independent of those effects. We conceptualize positive and negative wording valence effects as respectively reflecting subtle unconscious dispositions to *self-enhance* and *self-derogate*. Taken together those dispositions can be interpreted as "varnishing" more conscious forms of self-evaluation, and evaluation of one's future.[2]

1.1. Why include optimism in this study?

We include optimism in this study because items assessing this construct help us to estimate dispositions towards self-enhancement and self-derogation. Both of those dispositions are likely to influence responses to survey items used to assess evaluations of current self (i.e., self-esteem), and future self (i.e., dispositional optimism). Fontaine and Jones (1997) report a correlation of 0.80 between measures of

---

* Corresponding author at: Department of Sociology, University of Toronto, 725 Spadina Ave., Toronto, ON M5S 2J4, Canada.
  E-mail address: william.magee@utoronto.ca (W. Magee).
  URLs: https://www.researchgate.net/profile/William_Magee2/contributions; (W. Magee), https://www.researchgate.net/profile/Laura_Upenieks (L. Upenieks).

[1] We use the term valence in the same sense as emotion researchers (Charland & Gordon, 2006), as referring to positivity versus negativity of experience.

[2] Researchers have used different terms to refer to phenomena similar to self-derogation. For example, Owens (1993, 1994) refers to self-deprecation and self-depreciation. Others have referred to self-effacement. We prefer self-derogation because the other terms seem to connote action (e.g., speech), while we emphasize unconscious dispositions. See Humberg et al. (2019) for a recent review and discussion of labels and terms.

optimism and self-esteem, and Mäkikangas, Kinnunen, and Feldt (2004) report a correlation of 0.97 between latent factors of those constructs. This strong correlation observed between self-esteem and optimism is consistent with the idea that positivity underlies both self-esteem and dispositional optimism (Caprara, Eisenberg, & Alessandri, 2017), and with the measures of "core self-evaluation" (Arias & Arias, 2017; Gu, Wen, & Fan, 2015), which include items that assess current and future self. Wenglert and Rosén (1995) suggest that self-esteem can be conceptualized as encompassing aspects of optimism, such as the expectation that things that enhance one's self-worth will occur. The integration of optimism with self-esteem could reflect temporal self-integration, since people tend to perceive themselves over time as integrated entities (MacKinnon, 2015, pg. 40–41). Temporal integration of the self can bind the evaluation of one's current self to evaluations of one's future prospects, or possible selves (Markus & Nurius, 1986).

1.2. Interpreting wording valence effects distilled through bifactor models as reflecting dispositions to self-enhance and self-derogate

A number of previous studies have demonstrated wording valence effects in measures of self-esteem (Hyland, Boduszek, Dhingra, Shevlin, & Egan, 2014; Michaelides et al., 2016; Salerno et al., 2017), optimism (Vecchione, Alessandri, Caprara, & Tisak, 2014), and related constructs (McKay, Cole, & Percy, 2015). Research suggests that these wording valence effects can be substantial. For example, Alessandri, Caprara, and Tisak (2012) attributed 26% of variance in items on the commonly employed Rosenberg self-esteem scale to positive wording effects.

Some researchers have interpreted wording valence effects observed in studies of self-esteem and optimism in ways that suggest unconscious processes (Michaelides et al., 2016; Vecchione et al., 2014). For example, Vecchione et al. (2014) interpreted the latent factor for positive wording that they identified in an analysis of the Life Orientation Test-revised (LOT-R) as reflecting "self-deceptive enhancement." The notion of self-deception suggests that the effect of positive wording operates at an unconscious level, indicating an unconscious disposition to self-enhance that is distinguishable from a conscious tendency to expect that good things will happen in the future.

We build on the idea that wording effects reflect unconscious processes by drawing from research and theory on implicit measurement and implicit self-esteem (Falk & Heine, 2015; Greenwald & Banaji, 1995), and the idea that survey measures often assess both implicit and explicit cognitive processes (Fazio & Olson, 2003). We argue that self-esteem and optimism are both comprised, in part, by unconscious dispositions that are indicated by wording valence effects on agreement with self-related statements. Those effects can be interpreted as implicit measures of aspects of self-esteem and optimism.

In addition to expanding on Vecchione et al.'s (2014) interpretation of positive wording effects as reflecting unconscious self-enhancement, we draw on a number of literatures to argue that negative wording effects reflect an unconscious disposition to adopt a negative orientation towards current and future self. We label that disposition *self-derogatory* because it suggests readiness to accept negative statements about the self (Chang, Ferris, Johnson, Rosen, & Tan, 2012).

From the perspective of regulatory focus theory (Higgins, 1998), self-derogation, even if unconscious, might be adaptive if it increases preparedness to engage in appeasement behaviors, which in turn can maintain acceptance in groups (Keltner, 1995; Sznycer, Cosmides, & Tooby, 2017). Consistent with this, researchers have argued that self-derogation could reflect "a posture of conventional defense of individual worth" (Kaplan & Pokorny, 1969, pg. 425; see also Owens, 1993). For example, self-derogation could be consciously adopted as a strategy when individuals expect to be exposed to status degradation or stigmatization because they hold a typically devalued social identity characteristic, such as gender in some workplaces (Khan et al., 2017; Powell & Butterfield, 2015). We argue that self-derogation, as indicated by wording effects, reflects an unconscious defensive disposition.[3]

We refer to the measurement model that represents the idea that positive and negative wording valence effects reflect dispositions towards self-enhancement and self-derogation respectively as the unvarnishing bifactor (UB) model, presented in Fig. 1. Bifactor measurement models generally specify that observed responses reflect or "load onto" two latent factors (Marsh, Scalas, & Nagengast, 2010; Reise, 2012). The initial step in our analysis is to compare the fit of the UB model to data from the MIDUS study (Brim, Ryff, & Kessler, 2004) to the fit of a conventional measurement model in which only optimism and self-esteem factors are specified, represented in Fig. 2.

As we have already noted, the UB model posits that self-esteem and dispositional optimism are comprised of both unconscious tendencies to varnish evaluations and expectations along with less varnished evaluations and expectations about the future.[4] The idea that the self, and self-related attitudes such as self-esteem (MacKinnon, 2015) have multiple sub-dimensions or layers, at multiple levels of consciousness (Leary, Tambor, Terdal, & Downs, 1995; Major et al., 1999), has a long history in psychology (Roberts & Monroe, 1994). We refer to the layer captured by wording effects as *varnishes* because we think of the unconscious dispositions as lightening or darkening more deliberated aspects of self-evaluation, and evaluations of the future. We argue that the dispositions which constitute those varnishes are likely to be conditioned though interpersonal processes, including verbal feedback, reflected appraisals, and reinforcement (Deci, Olafsen, & Ryan, 2017; Felson, 2014).

The UB model can be understood as purging unconscious self-enhancement and self-derogation tendencies from optimism and self-esteem scores. The model also transforms the constructs of optimism and self-esteem into less "varnished" forms of each construct. For simplicity, we refer to these as "unvarnished" forms of self-evaluation and future orientation, although adjustment is made only for the varnishing that is associated with wording effects, not other forms of (more conscious or intentional) varnishing.

We conceptualize unvarnished future orientation (UFO) and unvarnished self-evaluation (USE) as reflecting judgments that stem from largely consciously accessible processes of deliberation about one's future and one's current self-sentiment. For UFO, the deliberation involves positive expectations. Yet, as the notion of "defensive pessimism" suggests (Norem & Cantor, 1986), even negative expectations about the future might serve an adaptive function. Men and women often have to adapt to different condition and expectations, and in the next section we outline reasons why we expect tendencies to self-derogate and self-enhance to vary by gender.

---

[3] One potential mechanism that might generate unconscious dispositions is suggested by polyvagal theory (Porges, 2011), which posits that the balance between defensive mobilization (i.e., fight, flight and immobilization) and social engagement is regulated at the level of vagus nerve activity. Self-defensive and self-enhancing responses at the neural level, are likely to have unconscious aspects. Unconscious aspects of polyvagal functioning could influence self-protection and self-enhancement orientation by shaping regulatory focus (Higgins, 1998). Evidence supporting the idea that unconscious processes underlie wording valence effects includes the finding that the positive wording valence effect is unrelated to social desirability (DiStefano & Motl, 2006). That finding suggests that thoughts and feelings about self that arise in interpersonal contexts, where people might react to reflected appraisals through conscious efforts to self-enhance (i.e., social desirability), are distinct from self-enhancement as indicated by positive wording effects.

[4] Conceptualizing wording effects in substantive terms does not imply that single factor scores, or raw scores obtained by summing observed responses to survey items assessing self-esteem or optimism are biased or invalid (Sartori & Pasini, 2007). We are not claiming that scores on those scales should necessarily be viewed as biased by wording effects. Instead, we are arguing that those constructs encompass dispositions and evaluations along multiple dimensions, including dimensions that are reflected in wording effects.
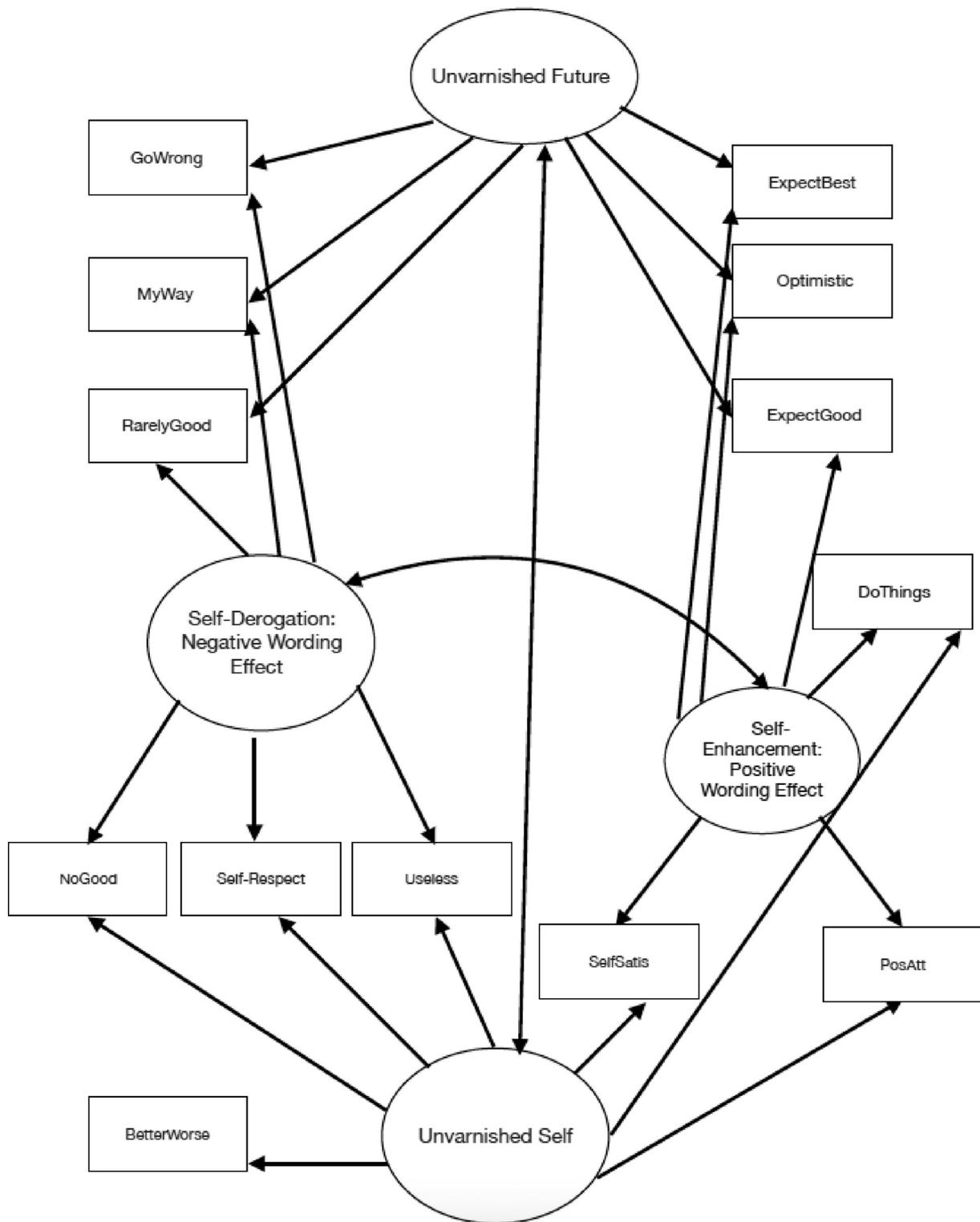
Fig. 1. The Unvarnishing Bifactor Model.
Note: See legend below Fig. 2.

1.3. The theoretical and substantive rationale for focusing on gender differences

As we previously noted, a number of prior studies have observed men to report slightly higher levels of self-esteem than women (Orth et al., 2010; Robins et al., 2002). One meta-analysis (Kling et al., 1999) estimated that the average gender difference in self-esteem across studies was about one-fifth of a standard deviation. We therefore expect to also observe only a small gender difference in self-esteem. However, we expect the dispositions to self-enhance and self-derogate to be more strongly gendered than is self-esteem as a whole. We also expect that the gender difference in the residual factor of unvarnished self-evaluation will be smaller than the gender difference in self-esteem.

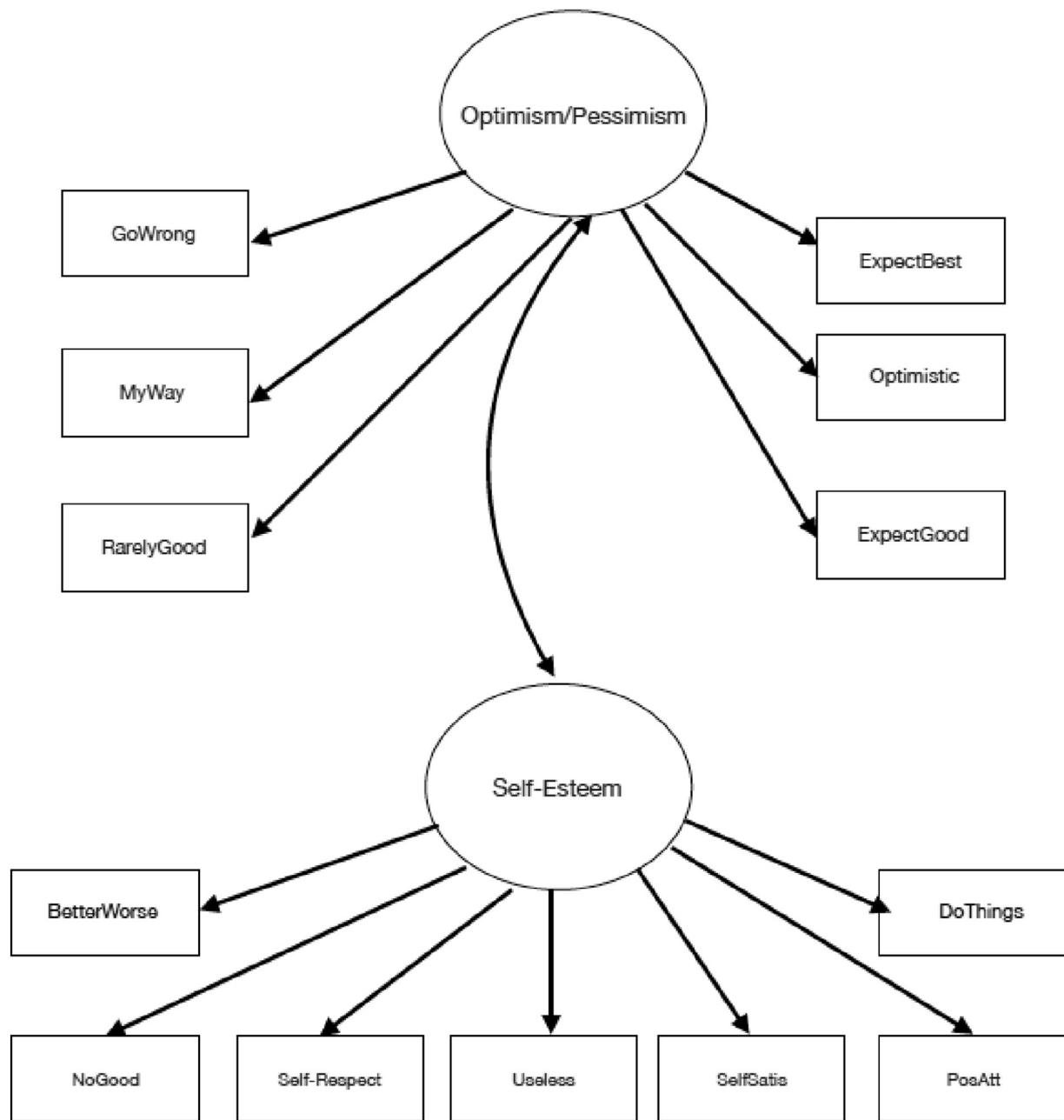Our expectations about gender differences are derived, in part, from

Fig. 2. Conventional Measurement Model
Note: Error terms are not shown in Fig. 1 and in this figure.

Summary of items for Figures 1 and 2.

Item	Description
GoWrong	If something can go wrong for me, it will
MyWay	I hardly ever expect things to go my way
RarelyGood	I rarely count on good things happening to me
ExpectBest	In uncertain times, I usually expect the best
Optimistic	I am optimistic about my future
ExpectGood	I expect more good things happen than bad
NoGood	I feel no good at all times
SelfRespect	I wish I had more respect for myself
Useless	I certainly feel useless at times
PosAtt	I take a positive attitude towards myself
DoThings	I am able to do things as well as most people
SelfSatis	On the whole, I am satisfied with myself
BetterWorse	I am no better/worse than others

regulatory focus theory (Higgins, 1998). That theory has motivated studies of both self-esteem (Leonardelli, Lakin, & Arkin, 2007) and future orientation (Zacher & de Lange, 2011), as well as other factors. Regulatory focus theory generally suggests that self-enhancement orientation, in both conscious and unconscious forms, stems from the possibility of envisioning success in one's attempts at self-promotion, or the promotion of one's interests.

Opportunity structures have historically been gendered in American society in ways that have entitled men (e.g., Blair-Loy, Hochschild, Pugh, Williams, & Hartmann, 2015; Budig, 2002), and we expect that male advantages and entitlements will be reflected in the development of a stronger habitual disposition among men than women to self-enhance. This disposition, as indicated by a tendency to agree with positively worded self-statements on surveys, is in turn likely to contribute to a gender difference in self-esteem.

A gender difference in self-derogation can also be interpreted from the perspective of regulatory focus theory. Here, self-derogation is understood as a defensive response that may prevent harm in the form of unanticipated derogation by others. One reason for expecting women to become more self-defensive than men is because they are exposed to many experiences that signal their relatively low valuation within organizations (Khan et al., 2017; Powell & Butterfield, 2015; Reuben, Rey-Biel, Sapienza, & Zingales, 2012). In addition, self-worth is partially continent on life experiences (Crocker & Wolfe, 2001), and some life experiences that are relevant to self-esteem are gendered (Kling et al., 1999) in ways that are likely to especially lower self-esteem among women. For example, studies have documented that women are not infrequently exposed to harassment (Bastomski & Smith, 2017; Gardner, 1995), incivility (Cortina et al., 2002; Kabat-Farr & Cortina, 2012), insults (Brinkman & Rickard, 2009; McCabe, 2009) and other forms of verbal degradation (Heilman, Wallen, Fuchs, & Tamkins, 2004). The literature on those exposures has been largely qualitative, focused on documenting the experience rather than estimating gender differences in rates of exposure. However, the literature does indirectly suggest that women are more likely to be exposed to self-derogation in the public sphere than men.[5]

The implications of status degradation for the development of self-defensive dispositions could be particularly great in contexts where high self-esteem is valued. In the United States self-esteem seems to be valued on par with the pursuit of happiness (Hewitt, 2002). High valuation of self-esteem and self-enhancement, in the context of gender inequality in status derogation, could shape gender differences in defensiveness, and unconscious dispositions that arise from repeatedly having to take a defensive stance. Thus, national differences in broad cultural valuations or ideology in conjunction with national differences in gender relations may be relevant to the diverse findings about gender difference in wording effects on self-esteem ratings. It is therefore notable that the three previous studies of gender differences in wording valence effects were each conducted in different national settings.

The earliest study of gender differences in scores on latent factors for wording valence effects analyzed data from college students in the American Southeast (DiStefano & Motl, 2009). That study estimated only one method factor –for negative wording. No mean gender difference for the negative wording method factor was observed in that study. Consistent with the literature on self-esteem in general, women in that study were found to score lower than men on the residual latent self-esteem factor.

A second study, by Lindwall et al. (2012), involved analyses of data from adult residents of a number of European countries, aged 60 years and older. They estimated a model that included both positive and negative wording effects, and observed no mean gender difference in

levels of either those factors. However, they found the structure of their measurement model to vary across national contexts, suggesting that different models might be necessary for different national populations.

The most recent study was conducted by Michaelides et al. (2016). That study, based on analyses of data from a German sample from the 1973 birth-cohort, included positive and negative wording factors in the measurement model, along with a residual factor labeled self-esteem. They observed no gender difference in summary scores for self-esteem, but women did score lower than men on the latent factor for obtained from their bifactor model (which they also referred to as self-esteem). No gender difference was observed for the negatively valenced wording factor, though there was a trend for women to also score higher on that factor. More importantly, contrary to our expectations, Michaelides et al. (2016) found scores on the positive wording factor to be higher for women than men. They suggested that the pattern of gender difference they observed in the wording factors might reflect a stronger polarization of emotions or emotion expression among women than men.

To reiterate, in contrast to Michaelides et al. (2016), we expect to observe a stronger positive wording valence effect among men than women. This expectation is based on our interpretation of positive wording effects as indicating self-enhancement orientation. We expect women to be more likely to agree with negatively worded statements about self (after wording effects are adjusted for shared variance in overall self-evaluation) because of women's greater likelihood of being exposed to experiences and conditions that should shape an automatic, unconscious tendency to self-derogate.

1.4. Age as a moderator of gender differences

In addition to evaluating hypotheses about effects of gender, we also consider whether gender differences in all outcomes vary with age. Previous research in the American context has found the gender difference in self-esteem to decrease in late life (Orth et al., 2010; Robins et al., 2002), after a period of growth through adolescence (Erol & Orth, 2011) or stability earlier in adulthood (Bleidorn et al., 2015). The early life trends suggest that gains in competence, achievement of goals, and increasing agency underlie increasingly positive self-related attitudes (see Deci et al., 2017, p. 27), including future orientation (Marques & Lopez, 2017). These age trends could reduce gender differences. However, although the gender difference in self-esteem seems to decline later in life, studies have observed the gender difference in self-esteem to remain at older ages (Orth & Robins, 2014). Research suggests that changes in optimism during the latter part of life shadow the trends in self-esteem (Chopik, Kim, & Smith, 2015; c.f. Glaesmer et al., 2012; Hinz et al., 2017; Schou-Bredal et al., 2017). We therefore evaluate the hypothesis that gender differences in the outcomes investigated here will decline with age by estimating interactions of age and gender.

2. Data and method

2.1. Sample

The data are from the second and third waves of the National Survey of Midlife Development in the United States (MIDUS) (Brim et al., 2004). MIDUS is a cohort longitudinal study. Adults were originally surveyed by telephone in 1995–1996 ($n = 7108$), with a response rate of 70%. Self-esteem was assessed by self-administered questionnaire only at the second and third waves of the study, conducted in 2004–2006 and 2013–2014 respectively. The estimated initial overall response rate was 60.8%, with 86.6% of the telephone respondents also completing the self-administered questionnaire. Consistent with previous studies conducted by researchers who were involved in the collection of the MIDUS data (e.g., Boehm, Chen, Williams, Ryff, & Kubzansky, 2015), we combine data from MIDUS samples, including

---

[5] For example, in a study about perceived "micro-aggressions" by McCabe (2009), men report events that indicate others are afraid of them, while women report blatant insults.

oversampled Milwaukee residents, with the data that were obtained through random-digit dialing of U.S. phone numbers.

The overall analytic sample excludes a small number of cases who did not respond to three or more of the self-esteem or optimism items (36 cases at wave 2, and 31 person-cases at wave 3). The final sample includes data from 4005 wave-2 cases, and 2699 wave-3 cases.[6]

2.2. Measures

2.2.1. Age and gender

Date of birth and gender are both self-reported in the telephone interview. Date of birth is used to determine age at date of interview. We use the term gender, rather than sex, because respondents were asked to report their gender presumably capturing self-perception (Tate, Ledbetter, & Youssef, 2013), albeit in dichotomous terms.

2.2.2. Self-esteem

Self-esteem was assessed by asking respondents to rate their level of agreement with seven self-related the statements using the five-point response scale of: "agree a lot", "agree a little", "neither agree not disagree"; "disagree a little", and "disagree a lot". The following statements are taken from the Rosenberg (1965) scale: "I feel no good at all, at times"; "I wish I could have more respect for myself"; "I certainly feel useless at times"; "I take positive attitude toward myself"; "I am able to do things as well as most people"; "On the whole, I'm satisfied with myself"; "I am no better and no worse than others". The first three items listed above are negatively worded, the next three are positively worded, and the last item is ambiguously worded. Given the ambiguity of the wording valence of the last-listed item, that item does not load onto either of the "wording" factors.

2.2.3. Optimism

The measure of optimism asks for level of agreement with self-referential statements, using the same response scale as the self-esteem measure. That measure is comprised of six items from the Life Orientation Test (LOT-R) (Carver & Scheier, 2014), three of which are positively worded and three that are negatively worded: "In uncertain times, usually expect best"; "I am optimistic about my future"; "I expect more good things happen than bad"; "If something can go wrong for me, it will"; "I hardly ever expect things to go my way"; "I rarely count on good things happen to me".

2.3. Estimating measurement models

The fit of measurement models to correlations among items is evaluated in the total person-wave data file, clustering on individuals. Correlation matrices of items (for men and women separately) are presented in Appendix A of supplementary materials available online. Estimates are obtained through full-information maximum likelihood (*mlmv* in Stata).

The models identify latent factors by fixing select item loadings to 1.0 (or −1.0 for self-derogation), resulting in metrics for most latent factors that mimic metrics for the item loadings, in order to aid interpretability. As noted in below, in Table 1, means for all latent factors are close to 4.0, which matches item means. The resulting range of values of the latent factors is slightly broader than the values of the raw scales, but the divergence in range is not large enough to compromise interpretability.

---

[6] We conducted parallel analyses using data from the most recent two waves of the *American's Changing Lives* (ACL) survey (collected in 2000–2001 and 2011–2012). Results of those analyses are available online (http…). We focus on the results from the MIDUS data because the ACL includes only three self-esteem items, rather than the seven contained in MIDUS. Analyses of the ACL data produce results that are consistent with those reported here.

Table 1
Descriptive statistics and sex differences.

	Men	Women		Total Sample
	(wave 2 n = 1790)	(wave 2 n = 2215)	Signif.	(wave 2 n = 4005)
	(wave 3 n = 1211)	(wave 3 n = 1488)	Sex	(wave 3 n = 2699)
	(either wave n = 2228)	(either wave n = 2814)	Diff.	(either wave n = 5042)
Age (wave2: 2004–2006)				
Mean	56.49	55.95		56.19
Std. dev.	12.24	12.52		12.39
Range	32–83	30–84		30–84
Age (wave 3: 2013–2014)				
Mean	64.72	64.27		64.48
Std. dev.	11.07	11.26		11.17
Range	42–92	39–93		39–93
Attitude scores: Conventional Model, both waves combined				
Optimism				
Mean	4.03	4.03		3.98
Std. dev.	0.75	0.81		0.80
Range	1.11–5.06	0.92–5.09		0.91–5.09
Self-esteem				
Mean	4.10	4.00	***	4.01
Std. dev.	0.92	0.92		0.93
Range	0.09–5.15	0.25–5.05		009–5.15
Attitude & disposition scores: Bifactor Model, both waves combined				
Unvarnished future				
Mean	4.04	3.99	*	3.98
Std. dev.	0.72	0.78		0.77
Range	1.36–5.52	0.96–5.63		0.96–5.63
Unvarnished self-evaluation				
Mean	4.05	4.03		4.01
Std. dev.	0.76	0.82		0.81
Range	0.41–6.00	0.35–6.10		0.35–6.10
Self-enhancement				
Mean	4.08	4.01	***	4.04
Std. dev.	0.53	0.60		0.58
Range	1.65–5.40	1.79–5.44		1.65–5.44
Self-derogation				
Mean	3.82	3.96	***	3.93
Std. dev.	0.80	0.83		0.83
range	1.73–6.97	1.74–7.41		1.73–7.41

* p < .05; ** p < .01; *** p < .001.

2.4. Estimating effects of gender and controls

Both formal tests of skewness (D'Agostino, Belanger, & D'Agostino Jr, 1990) and visualization of quantile plots indicate substantial divergence of all factors, except self-enhancement, from normality. Following Nichols (2010), who found that general linear models (GLMs) with a log-link provide more accurate estimates than alternative estimation procedures when outcomes are skewed, that estimation model is adopted. That model also takes the ordinal response scale of items into account. For self-enhancement, estimates from the Gaussian (i.e., normal) model are presented. Estimates based on the same procedures used in the other analyses (i.e. Poisson with log-link, available online, see supplementary materials) produce the same pattern of results.

GLMs are estimated for both cross-sectional analyses and mixed-effects multilevel general linear models (MEGLMs), specified with random intercepts for each individual. Models with and without a dummy indicator of wave of the MIDUS survey are estimated. Only the former are presented here, as inclusion of wave as a covariate does not substantially alter estimated effects of gender or age. The MEGLM results provide average effects over both waves, and permit tests of whether effects vary by wave. We also evaluate the interaction between gender and study wave in analyses of all outcomes to determine if

estimating average effects over both waves is reasonable. The effects of age and gender do not significantly vary by wave for any of the outcomes.

3. Results

3.1. Measurement model results

Comparison of model fit statistics (Akaike, 1987; Sclove, 1987) show the UB model to fit better than the conventional model, with the Bayesian Information Criteria (BIC) and Akaike Information Criteria (AIC) values for the UB model (e.g. BIC = 266,533.02; AIC = 266,172.1, df = 53) being smaller than those values for the conventional model (BIC = 270,367.3; AIC = 270,094.8, df = 40). According to all chi-square and residual based fit statistics (Bollen & Long, 1993), the UB model fits the data much better than the conventional model. Indeed, the UB model fits the data extremely well according to conventional criteria (e.g., CFI = 0.98; TLI = 0.98; RMSEA = 0.04). In contrast, the conventional model only fits the data moderately well (CFI = 0.86; TLI = 0.83; RMSEA = 0.10).

Wald and Score tests suggest statically significant gender differences in item loadings ($p < .05$) for both the conventional and the UB model.[7] Three item-loadings differ by gender for the UB model. Two of those paths are from UFO to the following two items: "I expect the best" and "I am optimistic". The third is the loading of the item, "I am no better or no worse than others" on the unvarnished self-evaluation factor. Since the specification of gender-specific loadings could work against finding gender differences in effects of latent factors, we estimated scores for factors where the loadings are specified as varying by gender, and for factors specified as gender invariant. Only results from the factor scores with gender-specific loadings are presented, though we found both age and gender effects to be consistent in analyses of gender-invariant and gender-variant factor scores. Previous studies of self-related attitudes that incorporate wording effects (DiStefano & Motl, 2009; Lindwall et al., 2012; Michaelides et al., 2016) have also found only minor gender variance in the factor structure. Indeed, we find that factor scores from models that specify gender-specific loadings are highly correlated ($r = 0.998–0.999$) with scores from gender-invariant models. Loadings, errors and latent factor correlations are presented in supplementary materials, Appendix A and B.

3.2. Distributions of demographics and attitudes

Information about the distribution of all variables is presented in Table 1. The mean values for each self-related attitude indicate that on average people agree "a little" with positive self-statements indicative of self-esteem, unvarnished self-evaluation, optimism, and unvarnished future orientations. Estimated mean level of agreement with self-derogation (negatively worded items) is slightly lower, but the variance and range of the self-derogation scores is slightly larger than for the other factors. Average self-derogation levels vary around a midpoint that can be interpreted as representing neither agreement or disagreement with negatively worded statements.

The baseline gender difference in observed self-esteem is small (about a tenth of a standard deviation), but as in previous research it meets the conventional ($p < .05$) criterion for statistical significance. There is no gender difference in observed optimism. Among the factors derived from the UB model, three of the four factors differ significantly by gender at baseline (i.e., with age not controlled). Although scores on UFO differ by only 0.05 points, that difference is significant at the $p < .05$ level. The baseline gender differences, for self-enhancement

and self-derogation, are larger, between one-tenth and one-fifth of a standard deviation ($p < .001$).

3.3. Estimated effects of gender and controls

Estimated effects of age and gender with optimism derived from the conventional two-factor measurement model are presented in the first column of coefficients in Table 2. These estimates are followed, to the right, by estimated effects on UFO scores derived from the unvarnishing bifactor (UB) model. Although the means and variances of the optimism and self-esteem factor scores are very similar to the means and variances for the UFO and USE score to which they are compared, there are small differences in scale ranges. Thus, z-scores are presented for comparative purposes.

Table 2 shows that that both optimism and unvarnished future orientation are similarly associated with age in a quadratic (inverted u-shaped) pattern. Consistent with previous research (Orth et al., 2010), preliminary analyses indicated that age is best modelled as having a quadratic effect on both self-esteem and optimism. That functional form extends to all outcomes, but is only of borderline statistical significance in the analysis of both waves of data combined, with id specified as a clustering variable.

In contrast to the baseline (i.e. bivariate) results for unvarnished future orientation presented in Table 1, the results presented Table 2 suggest that gender is not associated with UFO. This suggests that controlling for age, and taking skewness into account in the analytic approach, was sufficient to account for the gender difference.

In additional analyses of all outcomes, including UFO and optimism, we estimated the interaction between gender and age. The coefficients for those interactions did not meet conventional ($p < .05$) criteria statistical significance.

Table 3 presents results for self-esteem and unvarnished self-evaluation. Interactions with age and with the wave dummy variable were estimated to evaluate the stability of findings across waves. Those interactions were not statistically significant. Although the estimated effect of age on self-esteem is consistent with the effect of age on USE, gender effects on self-esteem and USE diverge. In the wave 2 analysis, self-esteem scores among men are significantly higher than among women (b = 0.027, SE = 0.007, $p < .05$). In contrast, unvarnished self-evaluation scores do not differ by gender (b = 0.007, SE = 0.006, $p > .05$).[8] In the wave-3 and combined analyses the gender effect on self-esteem is of borderline statistical significance, but in the analysis of USE it completely disappears, with a z-value of only 0.32.

Table 4 presents results from the analysis of self-enhancement. Results from both the single wave and combined analyses indicate that level of self-enhancement, indicated by the positive wording effect, is higher among men than women.

Table 5 presents results from the analysis of self-derogation. The age effects are the reverse of the pattern observed in the previous analyses of self-enhancement and unvarnished future orientation and self-esteem. However, unlike the self-enhancement results there is no overall difference in self-derogation across waves. This suggests that both the tendency for men to self-derogate less than women, and self-derogatory tendencies overall, seem to be more stable over waves than are self-enhancement tendencies.

4. Discussion

4.1. Putting the measurement model results in context

We find that the unvarnishing bifactor (UB) model, which includes latent factors for effects of positive and negative wording, fits the

---

[7] In the conventional model, three additional items have significantly different loadings on their factors by gender. These are: "can do", "satisfied with self" and "positive attitude".

[8] Supplementary analyses conducted with the ACL mirror these results. Gender is associated with self-esteem, but not USE.

Table 2
Estimated effects of age and gender on optimism & unvarnished perception of the future, from general linear models (Poisson: Log).

	Conventional: Optimism factor			Bifactor: Unvarnished future orientation		
	Wave 2	Wave 3	Both waves	Wave 2	Wave 3	Both waves
Age						
Coefficient	0.012***	0.020***	0.014**	0.010***	0.016***	0.011*
SE	0.002	0.035	0.005	0.002	0.003	0.005
z	5.55	5.65	2.77	4.41	4.67	2.18
Age-squared						
Coefficient	− 0.00009***	− 0.00014***	− 0.00010*	− 0.00007***	− 0.00010***	− 0.00008 [+]
SE	0.00002	0.00003	0.00004	0.00002	0.00000	0.00004
z	− 4.83	− 5.31	− 2.35	− 3.05	− 4.44	− 1.88
Male						
Coefficient	0.006	− 0.006	− 0.0021	− 0.009	− 0.017*	− 0.015
SE	0.006	0.007	0.0141	0.006	0.007	0.014
z	0.97	− 0.84	− 0.15	− 1.55	− 2.37	− 1.04
Wave 3 dummy						
Coefficient			− 0.029 [+]			− 0.025
SE			0.015			0.015
z			− 1.90			− 1.63
Constant						
Coefficient	0.979***	0.713***	0.937***	1.750***	0.86***	1.044***
SE	0.064	0.114	0.149	0.063	0.110	0.149
n	4005	2699	5042	4005	2699	5042

Notes: Random intercept specified for person (id) and as a clustering variable in MEGLM models.
*** $p < .001$.
** $p < .01$.
* $p < .05$.
[+] $p < .10$.

Table 3
Estimated effects of age and gender on self-esteem & unvarnished self-evaluation, from general linear models (Poisson: Log).

	Conventional: Self-esteem			Bifactor: Unvarnished self-evaluation		
	Wave 2	Wave 3	Both waves	Wave 2	Wave 3	Both waves
	Self-esteem	Self-esteem	Self-Esteem	Self-Evaluation	Self-Evaluation	Self-Evaluation
Age						
Coefficient	0.014***	0.022***	0.016**	0.009***	0.015***	0.012*
SE	0.003	0.004	0.005	0.002	0.004	0.005
z	5.32	5.27	3.28	3.91	4.36	2.35
Age-squared						
Coefficient	− 0.00010***	− 0.0015***	− 0.00011***	− 0.00006**	− 0.00011***	− 0.00008 [+]
SE	0.00002	0.00003	0.00004	0.00002	0.00003	0.00004
z	− 4.4400	− 4.8800	− 2.7000	− 3.0500	− 3.9700	− 1.8600
Male						
Coefficient	0.027***	0.016 [+]	0.02 [+]	0.007	− 0.002	0.005
SE	0.007	0.009	0.014	0.006	0.008	0.014
z	3.71	1.79	1.64	1.17	− 0.32	0.32
Wave 3						
Coefficient			− 0.0346**			− 0.033*
SE			0.01507			0.015
z			− 2.30			− 2.18
Constant						
Coefficient	0.913***	0.614***	0.832***	1.080***	0.848***	0.991***
SE	0.075	0.138	0.149	0.065	0.116	0.149
n	4055	2699	5042	4005	2699	5042

Notes: Random intercept specified for person (id) and as a clustering variable in MEGLM models.
*** $p < .001$.
** $p < .01$.
* $p < .05$.
[+] $p < .10$.

MIDUS data better than the conventional measurement model. Since previous studies have identified wording valence effects in measures of self-esteem and optimism (Hyland et al., 2014; Michaelides et al., 2016; Salerno et al., 2017; Vecchione et al., 2014), this result is not surprising. However, our interpretation of wording valence effects as representing "varnishings" associated with unconscious dispositions to self-enhance and self-derogate is novel. Although researchers have argued that wording effects "…might reflect a response style rather than an arti-fact…" (Lindwall et al., 2012, p. 202), and have argued that more effort should be made to interpret method effects in substantive terms (Reise, Moore, & Haviland, 2010), few researchers have attempted to provide substantive interpretations for gender differences in word-valence ef-fects. Before discussing the gender differences that emerged from our results, it is useful to briefly revisit findings from prior research that

Table 4
Estimated effects of age and gender on self-enhancement from general linear models (Gaussian, canonical).

	Wave 2	Wave 3	Both waves
Age			
Coefficient	0.015*	0.033**	0.016**
SE	0.007	0.010	0.005
z	2.240	3.220	3.160
Age-squared			
Coefficient	−0.000	−0.000**	−0.000*
SE	0.0000	0.0000	0.0000
z	−1.5300	−2.7800	−2.1800
Male			
Coefficient	0.085***	0.059**	0.067***
SE	0.018	0.022	0.020
z	4.72	2.72	3.35
Wave 3			
Coefficient			−0.084***
SE			0.013
z			−6.280
Constant			
Coefficient	3.470***	2.794***	3.416***
SE	0.184	0.331	0.156
n	4005	2699	5042

Notes: Random intercept specified for person (id) and as a clustering variable in MEGLM models.
*** p < .001.
** p < .01.
* p < .05.

Table 5
Estimated effects of age and gender on self-derogation from general linear models (Poisson: Log).

	Wave 2	Wave 3	Both waves
Age			
Coefficient	−0.013***	−0.017***	−0.012*
SE	0.002	0.004	0.005
z	−5.340	−4.520	2.450
Age-squared			
Coefficient	0.000***	0.000***	0.000*
SE	0.0000	0.0000	0.0000
z	4.8100	4.2600	2.2000
Male			
coefficient	−0.039***	−0.028***	−0.033*
SE	0.007	0.008	0.014
z	−5.770	−3.510	−2.320
Wave 3			
Coefficient			0.001
SE			0.015
z			0.090
Constant			
Coefficient	1.780***	1.937***	1.783***
SE	0.066	0.120	0.146
N	4005	2699	5042

Notes: Random intercept specified for person (id) and as a clustering variable in MEGLM models.
*** p < .001.
* p < .05.

provides context for our UB measurement model, and for how gender differences in levels of the factors in that model can be interpreted.

Previous research suggests that wording effects vary across populations. Lindwall et al. (2012) found the factor structure of measurement models that include wording valence effects to vary with national residential location of respondents within Europe. National differences in trends in self-esteem as conventionally defined and assessed have also been observed (Bleidorn et al., 2015). Geographic variation in the structure of self-esteem, and the dispositions, evaluations and sentiments that comprise it, suggest that cultural and structural factors might be at play. Cultural and structural differences between national

settings in gender relations, in conjunction with age differences in samples, might explain why our results, based on analyses of data from a U.S. population sample that encompasses a more than 50-year age-range, differs from the results of the Michaelides et al. (2016) study. Michaelides and colleagues analyzed a much smaller German sample of 38–39 year-old respondents.

4.2. A contextual interpretation of gender differences

As we noted, our results differ from what Michaelides et al. (2016) observed in a number of ways. First, consistent with prior research (Orth & Robins, 2014; Rentzsch et al., 2016; Robins et al., 2002), we observed a baseline gender difference in self-esteem, with men scoring slightly higher than women. In contrast, Michaelides et al. (2016) observed no baseline gender difference in self-esteem summary scores, a finding that is at odds with most previous research. Second, in contrast to our finding that men more strongly agree with the positively worded statements in survey assessments of self-esteem than women, Michaelides et al. (2016) observed women to more strongly endorse positively worded items than men.

In addition to the gender specificity in national differences noted above, this discrepancy between studies could be partially due to age differences in samples. All respondents in the Michaelides et al. (2016) study were aged 38–39 years old at the time of data collection. That is an age-rage when many people are engaged in childrearing. Childrearing is a highly gendered activity, and men and women could experience child rearing in ways that have implications for the gender difference in self-esteem. Consistent with this, research in the Netherlands suggests that the transition to parenting is associated with different patterns of self-esteem for men and women (Bleidorn et al., 2016). Thus, it could be that the Michaelides et al. findings especially reflect changes in dispositions associated with parenting or parenting-related transitions, among other age-related factors. Given the broader age-range of our sample, and our finding that age does not moderate gender effects, our results likely reflect stable gender differences among Americans.

Also in contrast to Michaelides et al. (2016), we find no gender difference in the factor that we call "unvarnished self-evaluation." Recall that unvarnished self-evaluation is identified by the inclusion of wording effects in the measurement model. Unvarnished self-evaluation is thus independent of wording valence effects, and the dispositions that underlie them. Our finding is consistent with the idea that the gender difference in self-esteem among Americans is largely due to gender differences in self-enhancing and self-derogatory dispositions that provide layers of "varnish" to self-evaluation.

A potential explanation for the gender differences that we observe in self-enhancement and self-derogation is that those differences are generated in part by exposure of American women to forms of treatment that lead many women to internalize sexism (Bearman, Korobov, & Thorne, 2009). Internalized sexism may be manifest at the level of unconscious response tendencies, such as the tendency to agree with self-derogatory statements. This raises the question of whether a woman exposed to sexist attitudes and behaviors, and who as a result develops and unconscious disposition to self-derogate, can at the same time hold a relatively positive unvarnished evaluation of herself. The lack of a gender difference in unvarnished self-evaluation scores suggests that this is possible.

If gender differences in exposure to harassment and other derogatory behaviors underlies our findings, it is useful to consider those practices in the context of the exercise of power in gender relations. Derogatory behaviors such as harassment might often reflect crude attempts to exercise power over women. However, power in gender relations is also exercised through more subtle discursive practices. Discursive practices associated with "benevolent sexism", such as patronizing language, have been found to shape task performance (Dardenne, Dumont, & Bollier, 2007). Future research might investigate

whether effects of subtle discursive practices extend to unconscious dispositions to self-derogate or self-enhance. If so, this might help explain way normative forms of discourse could have stronger effects than idiosyncratic (e.g., dyadic) personal communications (Felson, 2014).

Research has shown that subtle aspects of language are associated with self-protective and self-enhancement orientation (Semin, 2006). This supports our effort to theorize the development of dispositions in terms of process that could shape regulatory focus (Higgins, 1998). However, in emphasizing potential effects of discursive practices, we are not suggesting that effects of sexist discourse on dispositions to self-derogate or self-enhance must necessarily be maintained through elaborated internal speech. Wording effects probably occur too quickly to involve internal speech. Future research in this area might therefore investigate how dispositions, and regulatory focus orientations that presumably underlie them, could be shaped by discursive processes, but be maintained through non-linguistic processes.

4.3. A note about optimism and unvarnished future orientation

Although our emphasis in this study is on self-esteem and its dispositional subdimensions, the results with respect to optimism and unvarnished future orientation are worthy of note. The non-significant gender difference in optimism observed here is consistent with previous research (Boehm et al., 2015; Heinonen et al., 2006). Although we observed a small zero-order gender difference in unvarnished future orientation favoring men in Table 1, the gender differences observed in the regression analysis results for the total sample, presented in Table 2, did not meet the criterion for statistical significance. The possibility of a gender difference in unvarnished future orientation is intriguing enough, though, to encourage researchers to consider utilization of bi-factor models in the future to distinguish unvarnished forms of future-orientation from optimism.

Related to this, since unvarnished future orientation can be interpreted as being consonant with an agentic stance towards the future (Hitlin & Kirkpatrick Johnson, 2015), a potentially novel implication of research that builds on our approach might be that agency is sometimes undermined through unconscious processes that influence perceptions (i.e., perceptions of positively and negatively worded self-related statements), and immediate response tendencies (Strack & Deutsch, 2004). This possibility is important to consider because it would support the idea that raising consciousness about the potential effects of subtle aspects of language on the development of unconscious response dispositions could effectively counter dispositions that are associated with internalized sexism (Clegg, 1985).

4.4. A note about the lack of age difference in gender effects

Contrary to our expectations, we found no evidence that gender effects vary with age; none of the interactions between a dummy variable for gender and age was close to achieving statistically significance in any of the analyses, with *p*-values ranging from 0.15 to 0.97. However, one limitation of the current study with respect to age differences is that the MIDUS sample does not include adolescents and younger adults. Gender differences among young Americans, especially those in current youth cohorts (Twenge, Carter, & Campbell, 2017), might differ from the gender differences observed here. This would occur if processes generating gender differences in esteem, optimism, and dispositions to self-enhance or self-derogate change over time with changes in the relative status of women and men.

4.5. Some implications for related constructs & processes

Self-esteem and optimism are associated with other individual differences and forms of affect. Moreover, wording valence effects have been observed for a number of related constructs, including psychological distress and depression (Gu et al., 2015; Lindwall et al., 2012). We

therefore see possible extensions of our framework to related constructs and outcomes, especially those that are associated with regulatory focus, such as depression (Klenk, Strauman, & Higgins, 2011). Lindwall et al. (2012), found depression to be associated with a tendency to endorse negatively worded items. Since depression may be a reaction to the kinds of life experiences that we argue should promote self-derogation (e.g. harassment, derogation by others), and self-derogation could be both an antecedent of depression and a symptom of depression (Beck, 2002; Owens, 1994), that finding supports our emphasis on the experiential and social structural factors that might shape unconscious dispositions by shifting regulatory focus. It might therefore be profitable for future research to expand on our findings to investigate whether gender differences in wording valence effects in response to questions about depression, and other outcomes, are captured by the same factors that influence responses to self-esteem and optimism. If so, this would suggest that unconscious dispositions to self-enhance or self-derogate pervade self-assessment across constructs and dimensions of self. Evidence of this would provide warrant for much more research and theory development around wording effects.

Acknowledgement

Appendix A. Supplementary materials

Supplementary material relevant to this article can be found online at https://doi.org/10.1016/j.paid.2019.05.016.

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52,* 317–332.

Alessandri, G., Caprara, G. V., & Tisak, J. (2012). The unique contribution of positive orientation to optimal functioning. *European Psychologist, 17*(4), 44–54.

Arias, V. B., & Arias, R. (2017). The negative wording factor of core self-evaluations scale (CSES): Methodological artifact, or substantive specific variance? *Personality and Individual Differences, 109,* 28–34.

Bastomski, S., & Smith, P. (2017). Gender, fear, and public places: How negative encounters with strangers harm women. *Sex Roles, 76*(1–2), 73–88.

Bearman, S., Korobov, N., & Thorne, A. (2009). The fabric of internalized sexism. *Journal of Integrated Social Sciences, 1*(1), 10–47.

Beck, A. T. (2002). Cognitive models of depression. In R. L. Leahy, & T. E. Dowd (Vol. Eds.), *Clinical advances in cognitive psychotherapy: Theory and application. Vol. 14. Clinical advances in cognitive psychotherapy: Theory and application* (pp. 29–61). New York: Springer.

Blair-Loy, M., Hochschild, A., Pugh, A. J., Williams, J. C., & Hartmann, H. (2015). Stability and transformation in gender, work, and family: Insights from the second shift for the next quarter century. *Community, Work & Family, 18*(4), 435–454.

Bleidorn, W., Arslan, R. C., Denissen, J. J., Rentfrow, P. J., Gebauer, J. E., Potter, J., & Gosling, S. D. (2015). Age and gender differences in self-esteem—A cross-cultural window. *Journal of Personality and Social Psychology, 111*(3), 396–410.

Bleidorn, W., Buyukcan-Tetik, A., Schwaba, T., Van Scheppingen, M. A., Denissen, J. J. A., & Finkenauer, C. (2016). Stability and change in self-esteem during the transition to parenthood. *Social Psychological and Personality Science, 7*(6), 560–569.

Boehm, J. K., Chen, Y., Williams, D. R., Ryff, C., & Kubzansky, L. D. (2015). Unequally distributed psychological assets: Are there social disparities in optimism, life satisfaction, and positive affect? *PLoS One, 10*(2), e0118066.

Bollen, K. A., & Long, J. S. (Eds.). (1993). *Testing structural equation models.* Newbury Park, Calif: Sage Publications.

Brim, O. G., Ryff, C. D., & Kessler, R. C. (2004). *The MIDUS National Survey: An overview.* University of Chicago Press.

Brinkman, B. G., & Rickard, K. M. (2009). College students' descriptions of everyday gender prejudice. *Sex Roles, 61*(7–8), 461–475.

Budig, M. J. (2002). Male advantage and the gender composition of jobs: Who rides the glass escalator? *Social Problems, 49*(2), 258–377.

Caprara, G. V., Eisenberg, N., & Alessandri, G. (2017). Positivity: The dispositional basis of happiness. *Journal of Happiness Studies, 18,* 1–19.

Carver, C. S., & Scheier, M. F. (2014). Dispositional optimism. *Trends in Cognitive Sciences, 18*(6), 293–299.

Chang, C. H., Ferris, D. L., Johnson, R. E., Rosen, C. C., & Tan, J. A. (2012). Core self-evaluations: A review and evaluation of the literature. *Journal of Management, 38*(1), 81–128.

Charland, L. C., & Gordon, R. M. (2006). Emotion. In D. M. Borchert (Vol. Ed.),

Encyclopedia of philosophy. Vol. 2. Encyclopedia of philosophy (pp. 197–203). Detroit, MI: Macmillan Reference.

Chopik, W. J., Kim, E. S., & Smith, J. (2015). Changes in optimism are associated with Changes in Health over time among older adults. *Social Psychological and Personality Science, 6*(7), 814–822.

Clegg, S. (1985). Feminist methodology—Fact or fiction? *Quality and Quantity, 19*(1), 83–97.

Cortina, L. M., Lonsway, K. A., Magley, V. J., Freeman, L. V., Collinsworth, L. L., Hunter, M., & Fitzgerald, L. F. (2002). What's gender got to do with it? Incivility in the federal courts. *Law & Social Inquiry, 27*(2), 235–270.

Crocker, J., & Wolfe, C. T. (2001). Contingencies of self-worth. *Psychological Review, 108*(3), 593–623.

D'Agostino, R. B., Belanger, A., & D'Agostino, R. B., Jr. (1990). A suggestion for using powerful and informative tests of normality. *The American Statistician, 44*(4), 316–321.

Dardenne, B., Dumont, M., & Bollier, T. (2007). Insidious dangers of benevolent sexism: Consequences for women's performance. *Journal of Personality and Social Psychology, 93*(5), 764–779.

Deci, E. L., Olafsen, A. H., & Ryan, R. M. (2017). Self-determination theory in work organizations: The state of a science. *Annual Review of Organizational Psychology and Organizational Behavior, 4*, 19–43.

DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling, 13*(3), 440–464.

DiStefano, C., & Motl, R. W. (2009). Self-esteem and method effects associated with negatively worded items: Investigating factorial invariance by sex. *Structural Equation Modeling, 16*(1), 134–146.

Erol, R. Y., & Orth, U. (2011). Self-esteem development from age 14 to 30 years: A longitudinal study. *Journal of Personality and Social Psychology, 101*(3), 607–619.

Falk, C. F., & Heine, S. J. (2015). What is implicit self-esteem, and does it vary across cultures? *Personality and Social Psychology Review, 19*(2), 177–198.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*(1), 297–327.

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin, 116*(3), 429–456.

Felson, R. B. (2014). The (somewhat) social self: How others affect self-appraisals. In J. Suls (Vol. Ed.), *Psychological perspectives on the self. Vol. 4. Psychological perspectives on the self* (pp. 1–38). New York: Psychology Press.

Fontaine, K. R., & Jones, L. C. (1997). Self-esteem, optimism, and postpartum depression. *Journal of Clinical Psychology, 53*(1), 59–63.

Gardner, C. B. (1995). *Passing by: Gender and public harassment in public space.* Berkeley, CA: University of California Press.

Glaesmer, H., Rief, W., Martin, A., Mewes, R., Brähler, E., Zenger, M., & Hinz, A. (2012). Psychometric properties and population-based norms of the Life Orientation Test Revised (LOT-R). *British Journal of Health Psychology, 17*(2), 432–445.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*(1), 4–27.

Gu, H., Wen, Z., & Fan, X. (2015). The impact of wording effect on reliability and validity of the Core Self-Evaluation Scale (CSES): A bi-factor perspective. *Personality and Individual Differences, 83*, 142–147.

Heilman, M. E., Wallen, A. S., Fuchs, D., & Tamkins, M. M. (2004). Penalties for success: Reactions to women who succeed at male gender-typed tasks. *Journal of Applied Psychology, 89*(3), 416–427.

Heinonen, K., Räikkönen, K., Matthews, K. A., Scheier, M. F., Raitakari, O. T., Pulkki, L., & Keltikangas-Järvinen, L. (2006). Socioeconomic status in childhood and adulthood: Associations with dispositional optimism and pessimism over a 21-year follow-up. *Journal of Personality, 74*(4), 1111–1126.

Hewitt, J. P. (2002). The social construction of self-esteem. In C. R. Snyder, & S. J. Lopez (Eds.). *Handbook of positive psychology* (pp. 154–166). New York: Oxford University Press.

Higgins, E. T. (1998). Promotion and prevention: Regulatory focus as a motivational principle. *Advances in Experimental Social Psychology, 30*, 1–46.

Hinz, A., Sander, C., Glaesmer, H., Brähler, E., Zenger, M., Hilbert, A., & Kocalevent, R. D. (2017). Optimism and pessimism in the general population: Psychometric properties of the Life Orientation Test (LOT-R). *International Journal of Clinical and Health Psychology, 17*(2), 161–170.

Hitlin, S., & Kirkpatrick Johnson, M. (2015). Reconceptualizing agency within the life course: The power of looking ahead. *American Journal of Sociology, 120*(5), 1429–1472.

Humberg, S., Dufner, M., Schönbrodt, F. D., Geukes, K., Hutteman, R., Küfner, A. C. P., ... Back, M. D. (2019). Is accurate, positive, or inflated self-perception most advantageous for psychological adjustment? A competitive test of key hypotheses. *Journal of Personality and Social Psychology, 116*(5), 835–859.

Hyland, P., Boduszek, D., Dhingra, K., Shevlin, M., & Egan, A. (2014). A bifactor approach to modelling the Rosenberg Self Esteem Scale. *Personality and Individual Differences, 66*, 188–192.

Kabat-Farr, D., & Cortina, L. M. (2012). Selective incivility: Gender, race, and the discriminatory workplace. In D. Kabat-Farr, & L. M. Cortina (Eds.). *Gender and the dysfunctional workplace* (pp. 120–134). Northampton, MA: Edward Elgar Publishing.

Kaplan, H. B., & Pokorny, A. D. (1969). Self-derogation and psychosocial adjustment. *The Journal of Nervous and Mental Disease, 149*(5), 421–434.

Keltner, D. (1995). Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology, 68*(3), 441–454.

Khan, H. U. R., Khan, A., Zaman, K., Nabi, A. A., Hishan, S. S., & Islam, T. (2017). Gender discrimination in education, health, and labour market: A voice for equality. *Quality*

& Quantity, 51(5), 2245–2266.

Klenk, M. M., Strauman, T. J., & Higgins, E. T. (2011). Regulatory focus and anxiety: A self-regulatory model of Gad-Depression comorbidity. *Personality and Individual Differences, 50*(7), 935–943.

Kling, K. C., Hyde, J. S., Showers, C. J., & Buswell, B. N. (1999). Gender differences in self-esteem: A meta-analysis. *Psychological Bulletin, 125*(4), 470–500.

Leary, M. R., Tambor, E. S., Terdal, S. K., & Downs, D. L. (1995). Self-esteem as an interpersonal monitor: The sociometer hypothesis. *Journal of Personality Psychology, 68*, 518–530.

Leonardelli, G. J., Lakin, J. L., & Arkin, R. M. (2007). A regulatory focus model of self-evaluation. *Journal of Experimental Social Psychology, 43*(6), 1002–1009.

Lindwall, M., Barkoukis, V., Grano, C., Lucidi, F., Raudsepp, L., Liukkonen, J., & Thogersen-Ntoumani, C. (2012). Method effects: The problem with negatively versus positively keyed items. *Journal of Personality Assessment, 94*(2), 196–204.

MacKinnon, N. J. (2015). *Self esteem and beyond.* Palgrave Macmillan.

Major, B., Barr, L., Zubek, J., & Babey, S. H. (1999). Gender and self-esteem: A meta-analysis. In W. B. SwannJr., J. H. Langlois, & L. A. Gilbert (Eds.). *Sexism and stereotypes in modern society: The gender science of Janet Taylor Spence* (pp. 223–253). Washington, DC: American Psychological Association.

Mäkikangas, A., Kinnunen, U., & Feldt, T. (2004). Self-esteem, dispositional optimism, and health: Evidence from cross-lagged data on employees. *Journal of Research in Personality, 38*(6), 556–575.

Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist, 41*(9), 954–969.

Marques, S. C., & Lopez, S. J. (2017). In M. L. Wehmeyer, K. A. Shogren, T. D. Little, & S. J. Lopez (Eds.). *The development of hope*Springer Development of self-determination through the life-course.

Marsh, H. W., Scalas, L. F., & Nagengast, B. (2010). Longitudinal tests of competing factor structures for the Rosenberg Self-Esteem Scale: Traits, ephemeral artifacts, and stable response styles. *Psychological Assessment, 22*(2), 366–381.

McCabe, J. (2009). Racial and gender microaggressions on a predominantly-white campus: Experiences of black, Latina/o and white undergraduates. *Race, Gender & Class, 16*(1), 133–151.

McKay, M. T., Cole, J. C., & Percy, A. (2015). Further evidence for a bifactor solution for the consideration of future consequences scale: Measurement and conceptual implications. *Personality and Individual Differences, 83*, 219–222.

Michaelides, M. P., Zenger, M., Koutsogiorgi, C., Brähler, E., Stöbel-Richter, Y., & Berth, H. (2016). Personality correlates and gender invariance of wording effects in the German version of the Rosenberg Self-Esteem Scale. *Personality and Individual Differences, 97*, 13–18.

Nichols, A. (2010). *Regression for nonnegative skewed dependent variables.* (Paper presented at the BOS10 Stata Conference).

Norem, J. K., & Cantor, N. (1986). Defensive pessimism: Harnessing anxiety as motivation. *Journal of Personality and Social Psychology, 51*(6), 1208–1217.

Orth, U., & Robins, R. W. (2014). The development of self-esteem. *Current Directions in Psychological Science, 23*(5), 381–387.

Orth, U., Trzesniewski, K. H., & Robins, R. W. (2010). Self-esteem development from young adulthood to old age: A cohort-sequential longitudinal study. *Journal of Personality and Social Psychology, 98*(4), 645–658.

Owens, T. J. (1993). Accentuate the positive-and the negative: Rethinking the use of self-esteem, self-deprecation, and self-confidence. *Social Psychology Quarterly, 56*(4), 288–299.

Owens, T. J. (1994). Two dimensions of self-esteem: Reciprocal effects of positive self-worth and self-deprecation on adolescent problems. *American Sociological Review, 59*(3), 391–407.

Porges, S. W. (2011). *The polyvagal theory: Neurophysiological foundations of emotions, attachment, communication, and self-regulation (Norton Series on Interpersonal Neurobiology).* New York: WW Norton & Company.

Powell, G. N., & Butterfield, D. A. (2015). The glass ceiling: What have we learned 20 years on? *Journal of Organizational Effectiveness: People and Performance, 2*(4), 306–326.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*(5), 667–696.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*(6), 544–559.

Rentzsch, K., Wenzler, M. P., & Schütz, A. (2016). The structure of multidimensional self-esteem across age and gender. *Personality and Individual Differences, 88*, 139–147.

Reuben, E., Rey-Biel, P., Sapienza, P., & Zingales, L. (2012). The emergence of male leadership in competitive environments. *Journal of Economic Behavior & Organization, 83*(1), 111–117.

Roberts, J. E., & Monroe, S. M. (1994). A multidimensional model of self-esteem in depression. *Clinical Psychology Review, 14*(3), 161–181.

Robins, R. W., Trzesniewski, K. H., Tracy, J. L., Gosling, S. D., & Potter, J. (2002). Global self-esteem across the life span. *Psychology and Aging, 17*(3), 423–434.

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton, NJ: Princeton University Press.

Salerno, L., Ingoglia, S., & Coco, G. L. (2017). Competing factor structures of the Rosenberg Self-Esteem Scale (RSES) and its measurement invariance across clinical and non-clinical samples. *Personality and Individual Differences, 113*, 13–19.

Sartori, R., & Pasini, M. (2007). Quality and quantity in test validity: How can we be sure that psychological tests measure what they have to? *Quality & Quantity, 41*(3), 359–374.

Schou-Bredal, I., Heir, T., Skogstad, L., Bonsaksen, T., Lerdal, A., Grimholt, T., & Ekeberg, Ø. (2017). Population-based norms of the Life Orientation Test–Revised (LOT-R). *International Journal of Clinical and Health Psychology, 17*(3), 216–224.

Sclove, L. S. (1987). Application of model-selection criteria to some problems in

multivariate analysis. *Psychometrika, 52*(3), 333–343.

Semin, G. R. (2006). Modeling the architecture of linguistic behavior: Linguistic compositionality, automaticity, and control. *Psychological Inquiry, 17*(3), 246–255.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8*(3), 220–247.

Sznycer, D., Cosmides, L., & Tooby, J. (2017). Adaptationism carves emotions at their functional joints. *Psychological Inquiry, 28*(1), 56–62.

Tate, C. C., Ledbetter, J. N., & Youssef, C. P. (2013). A two-question method for assessing gender categories in the social and medical sciences. *Journal of Sex Research, 50*(8), 767–776.

Twenge, J. M., Carter, N. T., & Campbell, W. K. (2017). Age, time period, and birth cohort differences in self-esteem: Reexamining a cohort-sequential longitudinal study. *Journal of Personality and Social Psychology, 112*(5), e9–e17.

Vecchione, M., Alessandri, G., Caprara, G. V., & Tisak, J. (2014). Are method effects permanent or ephemeral in nature? The case of the Revised Life Orientation Test. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(1), 117–130.

Wenglert, Leif, & Rosén, Anne-Sofie (1995). Optimism, Self-Esteem, Mood and Subjective Health. *Personality and Individual Differences, 18*(5), 653–661.

Zacher, H., & de Lange, A. H. (2011). Relations between chronic regulatory focus and future time perspective: Results of a cross-lagged structural equation model. *Personality and Individual Differences, 50*(8), 1255–1260.