ORIGINAL ARTICLE

Measuring Appearance-Related Comparisons: Validation of the Comparison Standards Scale for Appearance

Nexhmedin Morina[1] · Peter McCarthy[1] · Thomas Meyer[1] · Pascal Schlechter[2]

Abstract

Background Humans constantly compare their attributes to different reference frames. According to the theoretical framework of the general comparative-processing model, such comparisons may be perceived as aversive (i.e., appraised as threatening the motives of the comparer) or appetitive (i.e., appraised as consonant with, or positively challenging the motives). However, we lack a measure that adequately captures multi-standard comparisons.

Methods Considering appearance-related comparisons as a relevant comparison domain, we introduce the Comparison Standards Scale for Appearance (CSS-A) that assesses appearance-related social, temporal, counterfactual, criteria-based, and dimensional upward and downward comparisons regarding their (a) frequency, (b) perceived discrepancy, and (c) engendered affect. We administered the CSS-A to 1121 participants, along with measures of appearance social comparison, body satisfaction, physical self-concept, self-esteem, well-being, and depression.

Results A two-factor model (*aversive and appetitive comparisons*) fit the data better than a bifactor model with an additional general domain-factor (*comparative thinking*). The validity of the CSS-A was supported by correlations with external validators beyond appearance, social comparison, and body satisfaction. Aversive comparisons displayed higher associations with most outcomes than appetitive comparisons.

Conclusions Overall, the CSS-A offers a psychometrically sound and useful measure of multi-standard comparisons.

Keywords Appearance · Body-image · Comparison theory · Well-being · Social comparison

Introduction

Features of the Comparison Process

Individuals constantly evaluate their mental or physical attributes that constitute the self. Comparison-based theories of judgment suggest that this evaluation lacks any utility scale and is based on frames of reference (Tversky, 1972; Vlaev et al., 2011). These theories posit that individuals make use of exemplars retrieved from memory or constructed through mental simulation to judge the value of the target in question (Morina, 2021; Stewart et al., 2006). Considering appearance as an example, the assumption is that people need a direct comparison with some standard to evaluate their appearance. Comparisons are best defined as a process comprising (and not limited to) (a) the selection of the comparison standard (e.g., social or temporal), (b) the basic comparison process of evaluating (dis-)similarities (e.g., between one's own appearance versus somebody else's) and producing the comparison outcome (i.e., the perceived discrepancy between the target and the standard), and (c) the engendered emotional, cognitive or behavioural responses (Morina, 2021). Comparisons of the target can occur against standards perceived as better off, worse off, as well as similar to the target (i.e., upward, downward, and lateral comparisons, respectively).

✉ Nexhmedin Morina
morina@uni-muenster.de

Peter McCarthy
mccarthy@uni-muenster.de

Thomas Meyer
t.meyer@uni-muenster.de

Pascal Schlechter
ps798@medschl.cam.ac.uk

[1] Institute of Psychology, University of Münster, Fliednerstr. 21, 48149 Münster, Germany

[2] Department of Psychiatry, University of Cambridge, Cambridge, UK

Crucially, the comparison outcome (i.e., the result of comparing the target with the standard) bears relevance to the individual's motives and goals. Consequently, comparisons can be defined as aversive (i.e., appraised as threatening the motives of the comparer), neutral, or appetitive (i.e., appraised as consonant with, or positively challenging the motives, Morina, 2021). The valuation of the comparison outcome then determines the comparer's reaction to the comparison outcome. So far, we lack empirical data about the extent to which aversive and appetitive comparisons represent two independent yet correlated factors. Alternatively, aversive and appetitive comparisons may represent components of an overarching latent factor accounting for a general comparison tendency.

Identifying and disentangling underlying mechanisms of comparison processes has important implications for research in psychology. However, surprisingly few studies have systematically investigated comparison processes in their entire breadth and complexity (Morina, 2021). This may be, in part, attributable to a lack of a measurement approach that adequately captures the (a) the frequency of upward and downward comparisons, (b) the degree of perceived discrepancy between the target and the standard, and (c) the engendered affective reactions. Drawing on a theory-driven approach, the present study therefore presents, refines, and validates the English version of the Comparison Standards Scale for Appearance (CSS-A), a scale that taps into (1) multiple comparison standards and (2) their motivational significance.

Types of Comparison and Their Motivational Significance

Several types of comparison that influence self-perception have been suggested in the literature: social (Festinger, 1954), temporal (Albert, 1977), counterfactual (Hoppen & Morina, 2021; Kahneman & Miller, 1986), criteria-based (Higgins, 1996; Lewin, 1951), and dimensional comparisons (Möller & Marsh, 2013). These comparison types share significant conceptual parallels and they all inform self-perception (Morina, 2021). Taking appearance as a target example, social comparisons involve comparing one's current appearance with someone else's. Temporal comparisons occur when one compares their current appearance with recollection of how they used to look at a certain time in the past or how they envision looking in the future. Counterfactual comparisons relate to comparing one's current appearance to that of a hypothetical self that might or should have occurred but did not actually occur and is thus counter to the facts. Criteria-based comparisons of one's current appearance occur against aspirations, norms, requirements, principles, or rules (e.g., how one ought to be looking at a certain age).

Finally, dimensional comparisons occur when one compares their current appearance with some other personal attribute.

Each comparison type can be broadly subdivided into upward, lateral, and downward comparisons, depending on the outcome of the comparison process. The comparison outcome is then valuated in terms of motives and coping. According to the general comparative-processing model (gComp, Morina, 2021), the motivational significance of comparison outcomes (i.e., aversive, neutral, or appetitive) varies between standards for upward vs. comparison outcomes. For example, perceiving myself as less good looking than my next-door neighbour (i.e., upward social comparison) and anticipating worse looks in the future (i.e., downward prospective temporal comparison) both represent aversive comparison outcomes. Conversely, downward social comparison (i.e., perceiving myself as better-looking than my neighbour) and upward prospective temporal comparison (i.e., anticipating better looks in the future) are both defined as appetitive comparison outcomes. Upward and downward dimensional comparison outcomes represent a special comparison type that also differs with respect to their motivational significance. Aforementioned comparison types distinguish between the target and the standard beyond current self-attributes, such as identity (e.g., my appearance relative to that of somebody else's) or time (e.g., my current appearance vs. my past appearance). In contrast, appearance-related dimensional comparisons are defined as thinking about one's own current appearance relative to other current personal attributes. As such, they comprise two or more personal attributes that differ only with respect to motivational significance (i.e., the comparer's valuation of each attribute). Generally, upward dimensional comparison outcomes may be defined as appetitive, in line with findings that upward dimensional comparisons increase positive affect (Möller & Husemann, 2006). Altogether, *upward* social, past temporal, counterfactual, and criteria-based comparisons, and *downward* prospective temporal comparisons can be defined as aversive (i.e., threatening the motives). On the other hand, *downward* social, past temporal, counterfactual, criteria-based, and dimensional comparisons, and *upward* prospective temporal and dimensional comparisons can be defined as appetitive outcomes (i.e., consonant with or challenging the motives).

Assessment of the Comparison Process

Scales that capture the complexity of the appearance-related comparison process are currently lacking. Given their shared conceptual parallels, we argue that to better understand the role of the comparison process in self-perception, different types of comparison, clustered into appetitive and aversive comparisons, need to be examined collectively. However, there are only domain-specific scales for appearance-based

social comparisons (O'Brien et al., 2009; Schaefer & Thompson, 2018; Thompson et al., 1991, 1999) and counterfactual thinking (Rye et al., 2008). The Physical Appearance Comparison Scale (Thompson et al., 1991) assesses the tendency to make personal physical appearance comparisons with others in various social situations. This scale was recently revised to additionally assess distal and proximal comparisons, upward and downward comparisons, and the emotional impact of comparisons (Schaefer & Thompson, 2018). Another instrument by Thompson et al. (1999), the Body Comparison Scale, assesses how often individuals compare specific body parts and more general body features to same sex peers. Finally, the Upward Physical Appearance Comparison Scale and Downward Appearance Comparison Scale (O'Brien et al., 2009) assess the frequency of upward and downward comparisons, respectively. Again, all these scales assess social comparisons only. Furthermore, they all assess perceptions of social comparisons in general and with regard to predefined situations (e.g., "*When I'm out in public, I compare my muscularity to the muscularity of others*" Schaefer & Thompson, 2018). Against this background, we aimed at developing a scale that a) assesses multiple types of comparison given the shared conceptual parallels across several comparison types and b) asks about the frequency of recent comparisons without predefining specific situations.

Social comparison represents by far the most prominent and widely studied type of comparison (Gerber et al., 2018), informing most of what we know about the role of comparison on self-perception. Overall, literature on social comparison suggests that individuals generally tend to choose an upward (rather than downward) comparison standard and to feel worse after an upward comparison and better after a downward comparison (Gerber et al., 2018). Other comparison types have been investigated to a much lesser degree but existing data demonstrate that temporal, counterfactual, criteria-based, and dimensional comparisons show parallels to social comparison and have similar effects on emotional and cognitive responses (Broomhall et al., 2017; Helm et al., 2017; Morina et al., 2022; Wilson & Shanahan, 2020). The propensity and strength, however, with which people habitually engage in multiple types of comparison and their clustering into aversive and appetitive comparisons have not yet been systematically investigated in concert and there exists no overarching approach to their measurement.

To address this research gap, we established a measure of individual differences in habitual comparison tendencies, using appearance-related comparisons as an example (McCarthy et al., submitted). We chose appearance for our purpose as appearance constitutes a salient construct for most individuals. Existing research on body (dis-)satisfaction has mostly focused on females (Tantleff-Dunn et al., 2011) and comparative findings suggest that women have somewhat higher levels of body dissatisfaction than men

(He et al., 2020). Research has further demonstrated that body dissatisfaction in both females and males is associated with low self-esteem, poorer psychological well-being, and depression (Barnes et al., 2020; Quittkat et al., 2019; Stice et al., 2000). Existing research on social comparison has demonstrated its crucial role in appearance, body image disturbance, and eating pathology (e.g., Hill & Nolan, 2021; Myers & Crowther, 2009). For example, patients with eating disorders reported a higher frequency of social comparisons than healthy control participants (Grynberg et al., 2020; Horndasch et al., 2015), as well as higher negative affect following social comparison (Vocks et al., 2010). Similarly, patients with body dysmorphic disorder reported a higher frequency of social comparison than healthy individuals (Anson et al., 2015). Although examined to a much lesser degree, other comparison types have also been significantly associated with well-being (Broomhall et al., 2017; Hoppen et al., 2020; Morina et al., 2022), suggesting that they play a significant role in appearance-related evaluations, reiterating the need to provide a solid assessment of multi-standard appearance related comparisons.

Against this background, we introduced the CSS-A to measure appearance-related comparisons for social, temporal, counterfactual, criteria-based, and dimensional standards. The CSS-A assesses (a) the frequency of upward and downward comparisons, (b) the degree of perceived discrepancy between the target and the standard, and (c) the engendered affective reactions. The assessment of these three components enables differential analyses on the consequences of the comparisons, thus providing a useful and flexible tool for researchers. In a first analysis of the German version of the scale (McCarthy et al., submitted), the frequency component was described by a bifactor model with an overarching latent factor and two orthogonal factors representing *upward* and *downward* comparisons. However, the discrepancy and affect components have not been psychometrically evaluated and it remains unknown whether they also adequately capture the underlying latent constructs. Furthermore, the psychometric properties of the English version of the scale have not been examined yet.

Present Study

In this study, we further refined and validated the CSS-A. In accordance with our main research question to examine whether aversive and appetitive comparisons represent two independent factors, we used confirmatory factor analyses (CFA) to test the factor structure of the CSS-A. Our approach was theory-driven (Morina, 2021) and based on initial findings with the German version of the scale (McCarthy et al., submitted). We hypothesised that all aversive (mostly upward) and all appetitive (mostly downward) comparisons would factor together, respectively. Note that we

expected the factor structure to be similar with regards to all three comparison components: frequency, discrepancy, and affect. We further expected aversive and appetitive factors to correlate positively with each other. To understand their shared variation better, we examined two different models that can be derived from the literature (Morina, 2021; McCarthy et al., submitted). In a first model, the two latent factors were expected to be correlated and to represent aversive and appetitive comparisons. In the second model, using a bi-factor approach, we introduced an overarching latent factor to account for a *general comparison tendency*. In addition to this factor, two orthogonal factors would account for the remaining variance in aversive and appetite comparisons. With respect to comparison frequency, the general factor would account for a general tendency to frequently compare one's appearance. With respect to comparison discrepancy, the general factor would account for a general tendency to perceive similar discrepancy between one's current appearance and the different comparison standards. Finally and related to engendered affect, the general factor would account for a general tendency to have similar affect upon engaging in appearance-related comparisons.

We hypothesised that comparison frequency and discrepancy will positively correlate with physical appearance social comparison and depressive symptoms and negatively with body satisfaction, physical self-concept, self-esteem, and psychological well-being. We further expected that the engendered affect will significantly negatively correlate with physical appearance social comparison and depressive symptoms and positively with body satisfaction, physical self-concept, self-esteem, and psychological well-being.

For a new scale to demonstrate validity, we deem it important to demonstrate that the CSS-A explains variance in the outcome variables after adjusting for related established constructs. Therefore, we expected that comparison frequency, discrepancy and the engendered affective impact will demonstrate incremental validity by predicting physical self-concept, self-esteem, psychological well-being, and depressive symptoms after adjusting for physical appearance social comparison and body satisfaction.

Method

Participants and Procedure

A total of 1121 study participants were recruited from online panel provider Prolific Researcher (Palan & Schitter, 2018). The sample size was based on recommendations that a sample size of more than 1000 participants implies lower measurement errors and more stable factor loadings (Boateng et al., 2018; Comrey & Lee, 2013). The survey was open to all panel members who had indicated to be fluent in English

and were older than 17 years. Participants were on average 28.7 ($SD = 9.7$) years old and 43.2% (n = 484) of them were female. Of the participants, $n = 202$ had a graduate degree, $n = 358$ a bachelor's degree, $n = 51$ an associate degree, $n = 229$ had some college education but no degree, $n = 268$ had a high school (or equivalent) degree, and $n = 13$ had no high school degree. The majority of participants ($n = 752$) was single or never married, followed by being married ($n = 345$). The study was approved by the Ethics Committee of the University of Münster. The survey material and the anonymized data can be found in the OSF supplement at https://osf.io/8sn5k/?view_only=8b74b4c3703f4e64bb95 412d163cd28b.

Measures

Comparison Standards Scale for Appearance (CSS-A): We recently developed the CSS-A to assess the degree of engagement in upward and downward comparisons via social temporal, counterfactual, criteria-based, and dimensional standards regarding one's own appearance. In line with the definition of comparison by Morina (2021) as considering comparative information in relation to the self to enable a judgment about relative standing, the CSS-A assesses comparison as thinking about one's own appearance in comparison to different standards. Both native English and German speakers were involved in the process of developing the scale, which was developed first in English and then translated to German. In several revision rounds, the final item pool was scrutinized regarding content coverage and clarity of language. During this process, items were refined whenever necessary. The German version was used for the first examination of the psychometric properties of the scale, which revealed that it can reliably and validly assess individual differences in the frequency of engagement in upward and downward comparisons (McCarthy et al., submitted). Following this examination and for parsimony reasons, we reduced the number of frequency items from 24 to 16. Of the remaining 16 items, eight are upward and eight are downward comparison items. The English version used in this study can be found in the OSF Materials supplementary folder (see above).

Table 1 provides a description of the comparison standards and the representative items as measured by the CSS-A. The scale comprises a) 16 obligatory items addressing frequency in the past 3 weeks on six-point Likert scales (0 = *not at all* to 5 = *very often*), b) 16 potential sub-items addressing discrepancy on a six-point Likert scale (0 = *not at all* to 5 = *much better/worse*), and c) 16 potential sub-items addressing the affective outcome on a bipolar seven-point Likert scale for affective impact (− 3 = *much worse* to + 3 = *much better*). For example, the upward past temporal comparison item first asks about the frequency "*Over*

Table 1 Descriptions for the frequency items of the CSS-A

Comparison Type	Standard (item #)	Over the *past 3 weeks* when considering your appearance, how often have you …
Social	Familiar (# 1/2)	Compared with others in your close circles who look *better/worse* than you
	Unfamiliar (# 3/4)	Compared with other individuals known and unknown to you who look *better/worse* than you
Temporal	Past (# 5/6)	Thought that you used to look *better/worse* than currently
	Future (# 7/8)	Thought that you might look *worse/better* in the future than currently
Counter-factual	Internal (# 9/10)	Thought that if you had behaved differently in the past, your appearance would now be *worse/better*
	External (# 11/12)	Thought that if others had behaved differently in the past, your appearance would now be *worse/better*
Criteria-based	Expectation (# 13/14)	Thought that you look *worse/better* than you had previously anticipated
Dimensional	Compensatory (# 15) (# 16)	Thought that *you have other personal attributes* that make up for what you lack in *appearance* Thought that *your appearance* makes up for what you lack in *other personal attributes*

[#]Refers to the item number in the CSS-A

the past 3 weeks when considering your appearance, how often have you thought that you used to look better than currently?". If participants indicate more than "0—not at all", they are asked "*How much better have you considered your past appearance to be?*" (i.e., discrepancy assessment) and "*On average during the past 3 weeks, how did the comparison make you feel?*" (i.e., affect assessment). In other words, participants only answered parts b) and c) of the respective item when they reported to have engaged in this comparison type. This way, we aimed to capture three relevant components of the construct while simultaneously being parsimonious with the number of questions. If participants indicated "0—*not at all*" to the frequency items, they received a score of zero on the respective discrepancy and affect items. This decision was based on the premise that individuals cannot be affected if they did not engage in any comparisons. This way, there were no missing data. We chose a 3-week recall period based on feedback from a pilot test with ten participants, who suggested that they would best recall comparisons made during this period.

Physical Appearance Comparison Scale (PACS; Thompson et al., 1991). The PACS was used to measure the tendency to make global social comparisons on the physical appearance domain (e.g., "*In social situations, I sometimes compare my figure to the figures of other people*"). This five-item scale asks participants to indicate the frequency with which they engage in five behaviours involving comparison with others in social settings. Higher PACS scores indicate a higher frequency of social comparisons on the appearance domain. The internal consistency for the current sample as indicated by Cronbach's alpha was 0.71.

Multidimensional Body-Self Relations Questionnaire—Appearance Evaluation Subscale (MBSRQ-AE; Brown et al., 1990). To assess body satisfaction, we used the seven-item MBSRQ-AE (e.g., "*I like my looks just the way they are*"). Items are rated on a 5-point Likert scale (1 = *definitely disagree*, 5 = *definitely agree*). Higher scores indicate greater body satisfaction. Cronbach's alpha in the current study was 0.91.

Multidimensional Self-Concept Scale—Physical Appearance Subscale (MSCS-P; Fleming & Courtney, 1984). The five-item Physical Appearance subscale of the MSCS was used to measure appearance-based self-concept. This subscale consists of five items (e.g., "*Have you ever felt ashamed of your physique or figure?*"). Participants are asked to rate each item on a 7-point Likert scale (1 = *not at all* or *never* to 7 = *very often* or *always*). A sum score of the items is used as an index of appearance-based self-concept, where higher scores indicate a more positive self-concept. Cronbach's alpha in the current study was .81.

Rosenberg Self-Esteem Scale (RSES; Rosenberg, 1965). The RSES was used to measure general self-esteem consisting of ten items. Items (e.g., "*On the whole, I am satisfied with myself*") are rated on a 4-point Likert scale (0 = *strongly disagree* to 3 = *strongly agree*). Higher RSES scores indicate a more positive self-esteem. Cronbach's alpha in the current study was .91.

Scales for Psychological Well-being (SPWB, Ryff & Keyes, 1995). To assess participants' level of well-being we used the 18-item SPWB that cover six areas of psychological well-being: autonomy, self-acceptance, environmental mastery, personal growth, positive relations with others, and

purpose in life. Items (e.g., *"I like most aspects of my personality"*) are rated on a 6-point scale (1 = *strongly disagree* to 6 = *strongly agree*). Higher SPWB scores are indicative of better well-being. In the current study, Cronbach's alpha was .87.

Patient Health Questionnaire (PHQ-8; Kroenke et al., 2009). Depressive symptoms were assessed with the eight-item PHQ-8 (e.g., "*Feeling tired or having little energy*"). The PHQ-8 assesses symptom severity over the last 2 weeks and is scored on a 4-point scale (0 = *not at all* to 3 = *nearly every day*). Higher PHQ-8 scores are indicative of more depressive symptoms. Cronbach's alpha in the current study was .88.

Data Analyses

Factor solutions. All analyses were conducted in R (R Core Team, 2019) version 4.01. The analysis code is openly available on the open science framework (see above). Items were treated as ordinal for all three comparison components: frequency, discrepancy, and affect. If participants indicated "0—*not at all*" to the frequency items, they received a score of zero on the respective discrepancy and affect items for our factor analyses to maximize available information. We used confirmatory factor analyses (CFA) to test our proposed models of the CSS-A (see Fig. 1). Two different theoretically derived models (McCarthy et al., submitted) were tested. First, we examined a two-factor solution with one factor representing *aversive* (mostly upward) comparisons and one factor representing *appetitive* (mostly downward) comparisons. To account for the fact that items of the same type of comparison (e.g., all social comparison or all temporal comparison items) may share common variance beyond the aversive and appetitive comparison factors, we allowed the covariances of errors for the same types of comparisons. Second, we tested a bifactor model with one overarching latent factor and two orthogonal specific factors. The latent overarching factor represents a *general comparison orientation* accounting for a general tendency of engaging in comparisons. The two specific factors represent the unique variance that is covered by an *aversive comparison* and an *appetitive comparison* factor. Both theoretically expected factor solutions were tested independently for all three comparison components (i.e., frequency, discrepancy, and affect). The *lavaan* package in R was used (Rosseel, 2012). As recommended for ordinal data, we used the weighed least squares mean and variance adjusted (WLSMV) estimator (Asparouhov & Muthén, 2010). Comparative Fit Index (CFI) and Tucker–Lewis Index (TLI) values > 0.95 indicate good fit and values > 0.90 indicate acceptable fit, root mean square error of approximation (RMSEA) and standardized root mean square residuals (SRMR) values < .05 indicate good fit and values < .08 indicate acceptable fit (Browne

& Cudeck, 1992; Hu & Bentler, 1999). Despite this theoretically grounded framework, we conducted additional exploratory factor analyses (EFA) separately for comparison frequency, discrepancy, and affect that we report in Supplemental Material 1. These EFA were conducted to ensure that we do not miss different relevant factor solutions of the CSS-A. As no further meaningful and consistent factor solution emerged in these analyses (see Supplemental Material 1 for a discussion), we continued with our a priori defined models in a CFA framework.

Measurement invariance across biological sexes. For a new measure, it is important to demonstrate that any differences between groups detected with the scale are unbiased. Given that there are sex differences in body satisfactions (He et al., 2020), we additionally tested whether the CSS-A measures the same underlying construct among males and females. To this end, we conducted measurement invariance analysis (MI), to discern whether the underlying construct is equally represented across males and females. (Meredith, 1993). In a multigroup CFA framework, increasingly constrained and hierarchically nested models were sequentially tested against each other. The respective constraints were added at each step in addition to the constraints introduced in the step before (Millsap, 2012). First, the factor structure was constrained to be equivalent across sexes (configural invariance). Second, the factor loadings were additionally constrained to be equal across sexes to discern whether the items relate to the proposed factors in the same way (weak/metric invariance). Third, item thresholds were additionally constrained to be equivalent to gauge whether the observed thresholds conditional on the latent factor do not differ across sexes (strong/scalar invariance). Fourth, the residual variances of the items were also constrained to be equal to scrutinize whether the amount of variance in the items not explained by the latent factors does not differ across sexes (strict/residual invariance; Meredith, 1993). Changes (Δ) in the CFI and RMSEA indicated violations of measurement invariance. The differences between the fit indices of two nested models suggest a violation of measurement invariance when ΔCFI exceeds .010 and ΔRMSEA exceeds .007 (Chen, 2007; Meredith, 1993).[1]

Nomological network. To examine the convergent validity of the CSS-A, we calculated scale composite scores for the *aversive comparisons* and the *appetitive comparisons* and correlated them with outcome measures that were theoretically expected to be associated with these two facets. In a next step, we aimed to demonstrate

---

[1] There is no clear consensus on these cut-offs for categorical data and we thus used rather strict cut-offs. The X^2-test statistics was not deemed suitable because it likely inflates minor differences if the sample size is sufficiently large.
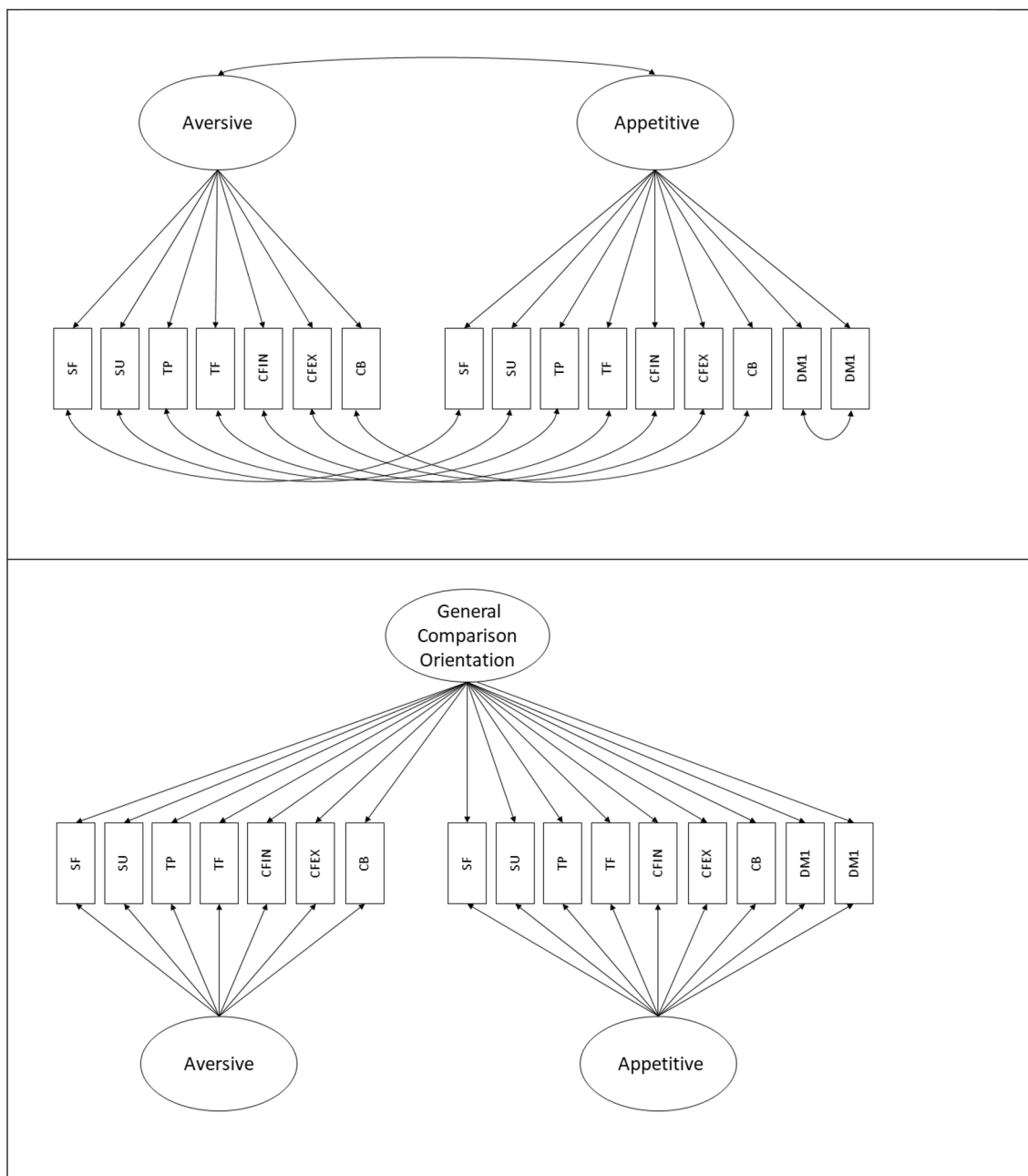
Fig. 1 Schematic representation of the two-factor solution with correlated errors (upper panel) and the bifactor solution (lower panel) *SF* social familiar, *SU* social unfamiliar, *TP* temporal past, *TF* temporal future, *CFIN* counterfactual internal, *CFEX* counterfactual external, *CB* Criteria-based, *DM* Dimensional

the incremental validity of the CSS-A beyond appearance-related social comparison and body satisfaction. To this end, we conducted multiple regression models using the two CSS-A subscales (*aversive* and *appetitive comparisons*) and adjusted for the tendency to compare one's appearance with others (i.e., PACS scores) and body satisfaction (i.e., MBSRQ-AE scores).

Results

Descriptive Statistics

All participants reported at least one appearance-related comparison. Aversive and appetitive comparisons were reported by 98.9% and 99.9% of participants, respectively.

Table 2 Descriptive statistics of the CSS-A facets

Item	Frequency					Discrepancy				Affect			
	N*	M	SD	Sk	Kt	M	SD	Sk	Kt	M	SD	Sk	Kt
1 (SF)	965	2.47	1.56	− 0.08	− 1.10	3.15	1.23	− 0.29	− 0.53	− 0.77	1.18	0.15	0.27
2 (SF)	806	1.56	1.37	0.55	− 0.67	2.06	0.97	0.10	0.00	0.16	1.03	− 0.10	0.94
3 (SU)	1007	2.87	1.51	− 0.44	− 0.76	3.48	1.19	− 0.52	− 0.30	− 0.93	1.25	0.22	− 0.11
4 (SU)	794	1.55	1.37	0.55	− 0.64	1.89	1.01	0.38	0.15	0.22	1.13	− 0.07	0.77
5 (TP)	923	2.58	1.69	− 0.19	− 1.20	3.14	1.34	− 0.42	− 0.47	− 0.85	1.41	0.32	− 0.23
6 (TP)	717	1.55	1.53	0.63	− 0.77	2.02	1.14	0.25	− 0.29	0.23	1.35	0.02	− 0.04
7 (TF)	775	1.96	1.70	0.30	− 1.23	1.94	1.19	0.40	− 0.20	− 0.87	1.35	0.32	0.04
8 (TF)	879	2.26	1.70	0.14	− 1.22	3.02	1.35	− 0.16	− 0.83	0.45	1.49	− 0.12	− 0.36
9 (CFIN)	868	2.62	1.83	− 0.20	− 1.37	3.38	1.25	− 0.48	− 0.41	− 0.88	1.36	0.48	0.25
10 (CFIN)	494	0.90	1.26	1.38	1.04	1.85	1.11	0.67	0.44	0.04	1.22	0.27	0.50
11 (CFEX)	390	0.90	1.43	1.39	0.64	2.77	1.25	0.10	− 0.62	− 0.42	1.25	0.14	0.54
12 (CFEX)	260	0.47	1.01	2.30	4.92	1.96	0.98	0.52	0.79	− 0.16	1.07	0.35	1.14
13 (CB)	722	1.68	1.59	0.47	− 1.01	1.93	1.04	0.32	0.09	− 0.96	1.19	0.13	− 0.10
14 (CB)	691	1.49	1.49	0.59	− 0.80	2.75	1.12	0.10	− 0.32	0.46	1.38	− 0.09	− 0.30
15 (DM)	872	2.39	1.69	− 0.08	− 1.21	3.21	1.13	− 0.28	− 0.30	0.67	1.40	− 0.20	− 0.23
16 (DM)	491	2.00	1.37	1.18	0.26	2.30	1.10	0.18	− 0.16	− 0.09	1.25	0.11	0.37

N* refers to the number of participants who indicated that they had compared at least once with the respective comparison standard. Discrepancy and affect values are based on these numbers

M mean, *SD* standard deviation, *Sk* skewness, *Kt* kurtosis, *SF* social familiar, *SU* social unfamiliar, *TP* temporal past, *TF* temporal future, *CFIN* counterfactual internal, *CFEX* counterfactual external, *CB* criteria-based, *DM* dimensional

With respect to types of comparison, dimensional comparisons were most frequently reported (99.6% of all participants), whereas criteria-based comparisons were reported the least (84.0% of all participants). Descriptive statistics for all three comparison components (i.e., frequency, discrepancy, and affect) and per item are depicted in Table 2. Note that this table depicts the descriptive statistics for the discrepancy and affect ratings of participants who indicated to have endorsed the respective comparison type (i.e., having at least a score of one on the frequency rating). Descriptive statistics for participants including added zeros for discrepancy and affect ratings for individuals who did not engage in the respective comparison type can be found in Supplemental Table 1. To have maximum power, the latter scores (with added zeros for the discrepancy and affect scores for participants who did not engage in the respective comparison type) were used for all subsequent analysis.

In terms of single items, the most frequently reported comparisons were Item 3 ("*compared with other individuals known and unknown to you who look better than you*"), Item 1 ("*compared with others in your close circles who look better than you*"), and Item 5 ("*thought that you used to look better than currently*"). Item 3 was also associated with the highest mean discrepancy between the target and the standard, followed by Item 9 ("*thought that if you had behaved differently in the*

past, your appearance would now be better") and Item 15 ("*thought that you have other personal attributes that make up for what you lack in appearance*"). With respect to affective impact, Item 13 ("*thought that you have other personal attributes that make up for what you lack in appearance*") was associated with the highest negative affect rating, followed by Items 3 and 9. Note, however, that with respect to affective impact all items but Item 12 ("*thought that if others had behaved differently in the past, your appearance would now be worse*") corresponded with the theoretical classification of aversive and appetitive comparisons. Item 12 that measures downward external locus of causation counterfactual comparison was associated with negative (rather than positive) affective impact.

Two-Factor Solution

Frequency

The model fit for all factor models can be found in Table 3. The two-factor model yielded good fit according to the CFI and acceptable fit according to the TLI, SRMR, and RMSEA. Table 4 displays all standardized factor loadings. All items loaded well on their respective factors (all λ > .40). Item 12 displayed the highest factor loading. The latent correlation between the two factors was *r* = .44.

Table 3 Model fit for the two-factorial solution and the bifactor model

	$\chi^2(df)$	p	CFI	RMSEA	SRMR	TLI
2 Factor model—Frequency	738 (92)	< .001	.951	.079	.075	0.936
Bifactor model—Frequency[a]	1242 (88)	< .001	.912	.108	.089	0.880
2 Factor model—Discrepancy	903 (92)	< .001	.915	.089	.081	0.889
Bifactor model—Discrepancy	682 (88)	< .001	.938	.078	.069	0.915
2 Factor model—Affect	236 (92)	< .001	.991	.037	.038	0.988
Bifactor model—Affect	498 (88)	< .001	.974	.065	.052	0.974

CFI Comparative Fit Index, *RMSEA* root mean square error of approximation, *SRMR* tandardized root mean square residuals, *TLI* Tucker–Lewis index, *df* degrees of freedom

[a]Some of the estimated variances were negative

Internal consistencies were also acceptable for the *aversive comparisons* factor ($\alpha = .73$; $\omega_{total} = .74$) and the *appetitive comparisons* factor ($\alpha = .73$; $\omega_{total} = .74$).

Discrepancy

The two-factor solution displayed acceptable fit according to the CFI and close to acceptable fit according to the TLI, SRMR, and RMSEA. Except for Item 15, all factor loadings for the *appetitive comparisons* factor were good (all $\lambda > .30$). Items 1, 3, and 5 had low factor loadings on the *aversive comparisons* factor. The latent correlation between aversive and appetitive comparisons was .72. Internal consistencies were acceptable for the *aversive comparisons* ($\alpha = .62$; $\omega_{total} = .63$) and the *appetitive comparisons* factor ($\alpha = .70$; $\omega_{total} = .71$).

Affect

All indices indicated good model fit for the two-factor solution. All items loaded acceptably on their respective factors (all $\lambda > .30$). A latent correlation of $r = .63$ emerged among the two factors. Internal consistencies were good for the *aversive comparisons factor* ($\alpha = .81$; $\omega_{total} = .81$) and acceptable for the *appetitive comparisons* factor ($\alpha = .71$; $\omega_{total} = .71$).

Bifactor Model Solution

Frequency

The bifactor model had acceptable fit according to the CFI and close to acceptable fit according to the other fit indices. However, some of the estimated variances were negative. Accordingly, the estimated parameters reported in Table 4 may be biased and hence need to be interpreted with caution.

Discrepancy

The bifactor model had acceptable model fit according to all fit indices. Some items loaded below a threshold of .30 on the *general comparison orientation* factor (items 1, 3, 5, and 9), yet had good factor loadings on the *aversive comparisons* factor. Item 15 had low factor loadings on the *general factor* and the *appetitive comparisons* factor.

Affect

The CFI and TLI indicated good model fit for the bifactor model solution, and RMSEA and SRMR indicated acceptable model fit for the bifactor model. All items displayed good factor loadings on the *general comparison orientation* factor (all $\lambda > .40$). All Items apart from Item 13 displayed good factor loadings on the *aversive comparisons* factor. On the *appetitive comparisons* factor only the two social comparison orientation items had good factor loadings (Items 2 and 4).

Model Comparison

Based on the results from the CFA analysis, we concluded that the two-factor solution has superior model fit compared to the bifactor model for comparison frequency and comparison affect. Despite somewhat better model fit of the bifactor model for comparison discrepancy, we proceeded with the two-factor solution for all comparison components in subsequent analyses and calculated respective means for overall scores per aversive and appetitive subscale for frequency (i.e., engagement in comparison), discrepancy (i.e., perceived [dis]similarity with the standard), and affective impact (i.e., engendered positive or negative affect). This was based on the premise that the scale should ideally capture the same factor structure for all three components (comparison frequency, comparison discrepancy and comparison affect). In this regard, the

Table 4 Factor loadings for the factorial solutions in the CFA

Item	Frequency					Discrepancy					Affect				
	2-Factors		Bifactor[a]			2-Factors		Bifactor			2-Factors		Bifactor		
	av	ap	General	av	ap	av	ap	General	av	ap	av	ap	General	av	ap
1 (SF)	.56	–	.61	.49	–	.27	–	.13	.76	–	.66	–	.42	.66	–
2 (SF)	–	.50	.31	–	.62	–	.45	.50	–	.46	–	.48	.52	–	.87
3 (SU)	.60	–	.62	.53	–	.24	–	.08	.82	–	.68	–	.37	.74	–
4 (SU)	–	.50	.32	–	.62	–	.45	.53	–	.62	–	.48	.49	–	.33
5 (TP)	.53	–	.41	.28	–	.28	–	.06	.37	–	.65	–	.43	.45	–
6 (TP)	–	.45	.09	–	.45	–	.53	.50	–	.06	–	.53	.52	–	.03
7 (TF)	.58	–	.47	.24	–	.54	–	.44	.10	–	.63	–	.49	.37	–
8 (TF)	–	.51	.15	–	.45	–	.37	.35	–	–.25	–	.48	.53	–	–.09
9 (CFIN)	.58	–	.47	.24	–	.41	–	.21	.48	–	.68	–	.43	.46	–
10 (CFIN)	–	.65	.28	–	.53	–	.71	.67	–	.04	–	.52	.49	–	.01
11 (CFEX)	.48	–	1.22	–1.29	–	.59	–	.61	.09	–	.47	–	.40	.47	–
12 (CFEX)	–	.70	.50	–	.51	–	.81	.90	–	–.06	–	.48	.49	–	–.07
13 (CB)	.66	–	.52	.29	–	.59	–	.45	.22	–	.67	–	.44	.21	–
14 (CB)	–	.55	.08	–	.59	–	.42	.42	–	–.30	–	.60	.58	–	.01
15 (DM)	–	.42	.22	–	.30	–	.26	.27	–	–.20	–	.38	.42	–	–.06
16 (DM)	–	.65	.28	–	.55	–	.63	.63	–	–.18	–	.41	.46	–	–.11

av aversive (mostly upward comparison direction factor), *ap* appetitive (mostly downward comparison direction factor), *general* general, *SF* social familiar, *SU* social unfamiliar, *TP* temporal past, *TF* temporal future, *CFIN* counterfactual internal, *CFEX* counterfactual external, *CB* criteria-based, *DM* dimensional

[a]Some of the estimated variance were negative. Hence, we do interpret the resulting parameter estimates

Table 5 Measurement invariance of the two-factorial solution between males (N = 625) and females (N = 484)

	χ^2(df)	CFI	RMSEA	ΔCFI	ΔRMSEA
2 Factor model—frequency					
Configural	857 (184)	.949	.081		
Metric	883 (198)	.948	.079	.001	.002
Scalar	1026 (260)	.942	.073	.006	.006
Strict	1074 (276)	.939	.072	.003	.001
2 Factor model—discrepancy					
Configural	1031 (184)	.915	.091		
Metric	1105 (198)	.909	.091	.006	.000
Scalar	1321 (260)	.891	.087	.018	.004
Scalarpartial	1270 (255)	.899	.084	.010	.007
Strictpartial	1337 (271)	.894	.084	.005	.000
2 Factor model—affect					
Configural	393 (184)	.986	.045		
Metric	437 (198)	.984	.047	.002	.002
Scalar	536 (260)	.983	.041	.001	.006
Strict	625 (276)	.978	.045	.005	.004

CFI Comparative Fit Index, *MSEA* root mean square error of approximation, *df* degrees of freedom, *scalar partial* we set the thresholds of item 3 free

two-factor structure presents the most consistent factor solution.

Measurement Invariance

Table 5 displays the measurement invariance analyses across female and male sex for the two-factorial solutions. We could establish the highest level of measurement invariance across sexes (strict invariance) for the frequency and affect components. Model fit did not deteriorate substantially for any of the tested components when increasingly constraining the model parameters. For the discrepancy component, the ΔCFI indicated a deterioration in model fit when comparing the metric invariance model with the scalar invariance model. Partial strict MI could be achieved, however, when setting the thresholds of Item 3 free. Females had lower thresholds to endorse the next response option of this item.

Nomological Network

Concurrent Associations

Table 6 shows single association of all three comparison components (frequency, discrepancy, and affect) and their respective two factors (aversive and appetitive comparison*s*), showing their significant intercorrelations. For most constructs, the *aversive comparisons factor* displayed descriptively higher correlations with the outcomes than the *appetitive comparisons factor.*

Frequency

A higher frequency of *aversive comparisons (M* = 2.15; *SD* = 1.00) correlated with all constructs in the expected direction, with male sex, and older age. On the other hand, a higher frequency of *appetitive comparisons (M* = 1.58; *SD* = 0.80) correlated positively with physical appearance social comparison, body satisfaction, self-esteem, overall psychological well-being, personal growth, and self-acceptance. It further correlated negatively with autonomy, male sex, and age.

Discrepancy

A higher discrepancy for *aversive comparisons (M* = 2.09; *SD* = 0.85) showed the same correlational patterns as the frequency subscale. A higher discrepancy for *appetitive comparisons (M* = 1.44; *SD* = 0.73) was positively associated with physical appearance social comparison, body satisfaction, physical self-concept, self-esteem, overall psychological well-being, and self-acceptance. This discrepancy was negatively associated with autonomy, and older age.

Affect

Recall that the affective impact with reference to all aversive and appetitive items of the CSS-A was assessed on a bipolar Likert scale from − 3 (*much worse*) to + 3 (*much better*). Results showed that less negative engendered affect for *aversive comparisons (M* = − 0.61; *SD* = 0.78) was positively associated with body satisfaction, physical self-concept, self-esteem, all aspects of psychological well-being (apart from purpose in life) and male sex. Less negative affect was further negatively related to physical appearance social comparison and depression. The same pattern emerged for more positive affect after engaging in *appetitive comparisons (M* = 0.17; *SD* = 0.54) with the difference that this was not associated with sex, but negatively correlated with purpose in life and older age (see Table 6).

Incremental Validity

Table 7 shows multiple regression models for the three comparison components (frequency, discrepancy, and affect), in which *aversive* and *appetitive* comparisons predict the outcome variables. To test the incremental validity of *aversive* and *appetitive* comparisons, we adjusted for physical appearance social comparison and body satisfaction in these models.

Table 6 Correlations with validators and among each other

	Frequency		Discrepancy		Affect	
	Aversive	Appetitive	Aversive	Appetitive	Aversive	Appetitive
Frequency—aversive	–	–	–	–	–	–
Frequency—appetitive	.29***	–	–	–	–	–
Discrepancy—aversive	.83***	.31***	–	–	–	–
Discrepancy—appetitive	.26***	.85***	.37***	–	–	–
Affect—aversive	− .56***	.26***	− .37***	− .29***	–	–
Affect—appetitive	− .22***	.34***	− .10***	.29***	.49***	–
PACS	.53***	.16***	.42***	.09***	− .39***	− .10***
MBSRQ-AE	− .45***	.24***	− .39***	.23***	.56***	.36***
MSCS-P	− .52***	.03	− .45**	.06*	.53***	.25***
RSES	− .48***	.11**	− .37***	.11***	.51***	.35***
PHQ-8	.48***	.05	.35***	.03	− .47***	− .26***
SPWB total	− .39***	.09***	− .32***	.07*	.36***	.35***
Autonomy	− .26***	− .10***	− .22***	− .13***	.11***	.17***
Environmental mastery	− .38***	.05	− .29***	.05	.39***	.25***
Personal growth	− .19***	.09***	− .17***	.06	.16***	.31***
Positive relations	− .25***	− .05	− .21***	.05	.24***	.20***
Purpose in life	− .13***	− .02	− .13**	− .03	.06	.16**
Self–acceptance	− .36***	.16***	− .21***	.15***	.40***	.35***
Male gender[a]	–3.13***	–1.05*	–2.19***	− 0.77	2.59***	0.49
Age	− 0.06*	− 0.25***	− .06*	− .24***	.00	− .07*

PACS Physical Appearance Comparison Scale, *MBSRQ-AE* Multidimensional Body-Self Relations Questionnaire—Appearance Evaluation Subscale, *MSCS-P* Multidimensional Self-Concept Scales—Physical appearance, *RSES* Rosenberg Self-Esteem Scale; *PHQ* Patient Health Questionnaire; *SPWB* Scales for Psychological Well-being

*p < .05, **p < .01, ***p < .001

[a]Estimated regression weight

Frequency

When adjusting for a physical appearance social comparison and body satisfaction, *aversive comparison* frequency was still associated with all constructs in the expected direction. *Appetitive comparison* frequency was positively associated with self-esteem, overall psychological well-being, and self-acceptance.

Discrepancy

After adjusting for physical appearance social comparison and body relations *aversive comparison* discrepancy was associated with all constructs except for autonomy and personal growth. *Appetitive comparison* discrepancy was negatively related to autonomy, and positively associated with self-esteem and self-acceptance.

Affect

When adjusting for physical appearance social comparison and body satisfaction, less negative engendered affect in the context of *aversive comparison*s was positively associated with all measured constructs apart from overall psychological well-being and self-acceptance. More positive engendered affect in the context of *appetitive comparisons* was positively associated with self-esteem, overall psychological well-being, autonomy, environmental mastery, personal growth, purpose in life and self-acceptance.

Discussion

The CSS-A reliably and validly assesses between-person differences in five types of appearance-related comparison. All participants indicated engaging in some sort of comparison, with both aversive and appetitive comparisons being endorsed by nearly all participants. The most frequently endorsed type of comparison standard was dimensional comparison, however even the least frequently reported type of comparison standard (criteria-based comparison) was reported by 84% of participants. Comparison frequency, discrepancy, and affective impact were significantly

Table 7 Multiple regression weights when adjusting for PACS and MBSRQ-AE

	Frequency				Discrepancy				Affect			
	β	SE	p	R^2	β	SE	p	R^2	β	SE	p	R^2
MSCS-P												
CSS-A—av	−0.18	0.03	< .001	.53	−0.18	0.03	< .001	.53	0.21	0.04	< .001	.52
CSS-A—ap	0.01	0.03	.944		0.02	0.03	.416		−0.04	0.04	.254	
PACS	−0.41	0.04	< .001		−0.46	0.04	< .001		0.47	0.03	< .001	
MBSRQ-AE	0.54	0.03	< .001		0.55	0.03	< .001		−0.53	0.04	< .001	
RSES												
CSS-A—av	−0.25	0.03	< .001	.41	−0.16	0.03	< .001	.38	0.25	0.04	< .001	.41
CSS-A—ap	0.10	0.03	< .001		0.07	0.03	.007		0.10	0.04	< .001	
PACS	−0.19	0.04	< .001		−0.27	0.04	< .001		−0.19	0.03	< .001	
MBSRQ-AE	0.38	0.03	< .001		0.44	0.03	< .001		0.38	0.04	< .001	
PHQ-8												
CSS-A—av	0.31	0.03	< .001	.27	0.18	0.03	< .001	.21	−0.34	0.04	< .001	.25
CSS-A—ap	−0.02	0.03	.489		0.01	0.03	.755		0.04	0.04	.230	
PACS	0.07	0.04	.130		0.21	0.04	< .001		0.18	0.03	< .001	
MBSRQ-AE	−0.18	0.03	< .001		−0.25	0.03	< .001		−0.14	0.04	< .001	
SPWB overall												
CSS-A—av	−0.57	0.08	< .001	.27	−0.42	0.09	< .001	.25	0.02	0.01	.882	.28
CSS-A—ap	0.21	0.07	.003		0.11	0.08	.170		0.72	0.10	< .001	
PACS	−0.26	0.13	.048		−0.42	0.12	< .001		−0.59	0.12	< .001	
MBSRQ-AE	0.86	0.09	< .001		0.99	0.08	< .001		0.92	0.08	< .001	
Autonomy												
CSS-A—av	−0.07	0.02	< .001	.08	−0.04	0.02	.059	.09	−0.07	0.02	.003	.09
CSS-A—ap	−0.03	0.02	.085		−0.06	0.02	< .001		0.12	0.02	< .001	
PACS	−0.14	0.03	< .001		−0.14	0.03	< .001		−0.21	0.03	< .001	
MBSRQ-AE	0.07	0.02	< .001		0.06	0.02	.002		0.05	0.02	.008	
Environmental mastery												
CSS-A—av	−0.14	0.02	< .001	.22	−0.08	0.02	< .001	.20	0.11	0.02	< .001	.22
CSS-A—ap	0.03	0.02	.053		0.02	0.02	.407		0.05	0.02	.045	
PACS	−0.05	0.03	.119		−0.10	0.03	.001		−0.09	0.03	.001	
MBSRQ-AE	0.17	0.02	< .001		0.23	0.02	< .001		0.17	0.02	< .001	
Personal Growth												
CSS-A—av	0.04	0.02	.026	.10	−0.04	0.02	.088	.10	−0.11	0.02	< .001	.16
CSS-A—ap	0.03	0.02	.087		0.01	0.02	.530		0.20	0.03	< .001	
PACS	0.01	0.03	.350		−0.01	0.03	.621		−0.07	0.02	.001	
MBSRQ-AE	0.14	0.02	< .001		0.15	0.03	< .001		0.15	0.03	< .001	
Positive relations with others												
CSS-A—av	−0.10	0.02	< .001	.10	−0.09	0.03	< .001	.09	0.07	0.02	.008	.10
CSS-A—ap	0.03	0.02	.078		−0.03	0.02	.109		0.05	0.02	.138	
PACS	−0.03	0.02	.763		−0.05	0.03	.104		−0.06	0.02	< .001	
MBSRQ-AE	0.11	0.03	< .001		0.15	0.02	< .001		0.11	0.03	.062	
Purpose in life												
CSS-A—av	−0.05	0.02	.024	.03	−0.04	0.02	.008	.03	−0.06	0.03	.019	.03
CSS-A—ap	−0.00	0.02	.956		−0.02	0.02	.264		0.07	0.03	.003	
PACS	−0.02	0.02	.541		0.00	0.03	.747		0.04	0.02	.189	
MBSRQ-AE	0.06	0.03	< .001		0.07	0.02	< .001		0.08	0.03	< .001	
Self-acceptance												
CSS-A—av	−0.13	0.02	< .001	.28	−0.09	0.02	< .001	.28	0.06	0.03	.069	.30
CSS-A—ap	0.09	0.02	< .001		0.07	0.02	< .001		0.16	0.03	< .001	
PACS	−0.09	0.02	.008		−0.12	0.03	< .001		−0.13	0.02	< .001	
MBSRQ-AE	0.25	0.03	< .001		0.28	0.02	< .001		0.26	0.03	< .001	

PACS Physical Appearance Comparison Scale, *MBSRQ-AE* Multidimensional Body-Self Relations Questionnaire—Appearance Evaluation Subscale, *MSCS-P* Multidimensional Self-Concept Scales—Physical appearance, *RSES* Rosenberg Self-Esteem Scale, *PHQ* Patient Health Questionnaire; *SPWB* Scales for Psychological Well-being; *SE* standard error, *av* aversive (mostly upward comparison direction factor), *ap* appetitive (mostly downward comparison direction factor

intercorrelated. The findings further suggest that the structure of the CSS-A is primarily characterized by an aversive factor and an appetitive factor. Our findings additionally support a two-factor structure and provide evidence of convergent and incremental validity.

Factor Structure

The two-factor model fit the data better than the bifactor model. The items within the aversive and appetitive factors had significant loadings on their respective factors, while in the bifactor model there were many noticeably stronger associations on the specific factors than the general factor. This suggests that the aversive and appetitive constructs, both capture unique variance of comparison scores. In our previous study with a German sample (McCarthy et al., submitted), we conducted the factor analysis only with the frequency items of the CSS-A. The examination of the bifactor model in the previous study supported this model at large. However, the model first needed adjustment by removing three items that were non-significant indicators of the global comparison factor (one upward past temporal comparison item, one downward past temporal comparison item, and one downward criteria-based comparison item). With this adjustment, the SRMR was still below the threshold of an acceptable fit (0.103), whereas the other fit indices were satisfactory. Our previous study was, however, based on a much smaller sample (n = 300) and did not examine a two-factor solution only. The present study with a much larger sample indicates that the two-factor solution is a better fit to the data than the bifactor model. In the present study, we also allowed for the covariances of error between items of the same comparison type (e.g., all social comparison items), based on the assumption that they share common variance beyond their respective factors.

The classification in aversive and appetitive comparisons was based on the theoretical motivational valence of comparison processes (Morina, 2021). Accordingly, it was expected that all comparisons theoretically defined as aversive (vs. appetitive) would be associated with negative (vs. positive) affect. This did not apply for Item 12 ("*thought that if others had behaved differently in the past, your appearance would now be worse*"), however, which was associated with negative (rather than positive) affect. This suggests that thinking that one's appearance would now be worse if others had behaved differently in the past triggers rather negative emotions. We can only speculate as to why this is. The exact wording of the CSS-A item on affective impact is "*how did the comparison make you feel?*" and it is likely that study participants reported how they felt following the comparison irrespective of what triggered negative emotions. Thinking what might be different now if others had behaved differently in the past may have also triggered aversive memories

of how others have treated the comparer badly, with the consequence that such memories were then associated with negative emotions. Experimental studies are needed to accurately examine the association between the direction of appearance-related comparisons (i.e., counterfactual and otherwise) and engendered affective responses.

Measurement Invariance

For the frequency and affect scale, we could establish strict measurement invariance across sexes. For the discrepancy subscale, partial strict measurement invariance was found after we allowed the thresholds of one item to freely vary across groups. The specific item was: "…*compared with other individuals known and unknown to you who look better than you?*", with the follow-up discrepancy question: "*How much better have you considered their appearance to be?*". One explanation relates to the result that females reported higher discrepancy between their current appearance and the appearance of upward unfamiliar others than males. This is in line with findings that females have somewhat higher levels of body dissatisfaction than men (He et al., 2020), and our data indicate that this is particularly pronounced in relation to unfamiliar social comparisons. However, this assumption needs to be accurately examined in future research.

The overall high level of measurement invariance is an important feature of our scale because appearance is a topic where sex differences are expected to emerge (He et al., 2020). However, between-group differences can also be attributable to differences in the measurement properties of the scale when the scale is biased. Establishing measurement invariance allows for the interpretation that scores are unbiased across sexes and that potential differences can be attributed to actual difference in the latent construct (Meredith, 1993). Accordingly, sex differences detected with the CSS-A can be meaningfully interpreted. Females reported a higher mean frequency of aversive and appetitive comparisons. Further, they reported higher discrepancies when engaging in aversive comparisons. Males on the other hand reported less negative affect in the context of aversive comparisons. These sex differences indicate that females are more likely to engage in appearance-based comparisons with stronger emphasis on differences, as well to experience more negative affect upon aversive comparisons than males.

Validity

Our findings further yielded that the aversive and appetitive subscales of the CSS-A were significantly correlated with physical appearance social comparison, body satisfaction, physical self-concept, self-esteem, psychological well-being, and depressive symptoms. Importantly, this applied to all

measured components of comparison, i.e., frequency, discrepancy, and affective impact. This suggests that all three components are suitable to be applied to examine different research questions related to self-evaluations. Yet, overall aversive comparison displayed descriptively higher correlations with the measured outcomes than appetitive comparisons. Aversive and appetitive comparisons proved also to be significant predictors of physical self-concept, body satisfaction, self-esteem, depression, and most facets of psychological well-being over and beyond physical appearance social comparison and body satisfaction. However, here too, aversive comparisons showed stronger incremental validity. As such, the findings of convergent and incremental validity further speak to the usefulness of separating aversive comparisons from appetitive ones. We need to consider, however, that the relevance of appetitive scores may be context-sensitive. Therefore, future research needs to investigate the comparative role of aversive and appetitive comparisons in different self-relevant dimensions, such as physical or psychological well-being or performance. Altogether, being able to conduct analyses on both forms of comparisons (i.e., aversive vs. appetitive) for three different comparison components (frequency, discrepancy, and affect) provides a comprehensive and flexible application of the CSS-A in different research contexts related to self-evaluation.

Previous assessments of habitual comparison standards have examined unitary aspects of the comparison process and mostly the frequency of specific types of comparison, such as social comparison (Allan & Gilbert, 1995; Schaefer & Thompson, 2018; Thompson et al., 1991) and to a lesser degree counterfactual thinking (Rye et al., 2008). With respect to body (dis-)satisfaction and eating pathology, previous research has revealed significant insights into the role of social comparison therein (Hill & Nolan, 2021; Myers & Crowther, 2009). Our results are in line with findings on social comparison showing that individuals generally tend to feel worse after an upward social comparison and better after a downward social comparison (Gerber et al., 2018). Yet, there is lack of research on the role of other types of comparison in body (dis-)satisfaction. Considering various comparison types and standards, comparison direction, perceived (dis)similarity, as well as engendered affective, cognitive, and behavioural reactions will most likely increase our knowledge about comparison processes altogether as well as their role in psychopathology (Gerber et al., 2018; McCarthy & Morina, 2020; Morina, 2021; Wood, 1996). Our approach is based on this hypothesis and the present results suggest that the CSS-A offers a reliable and valid tool to assess key components of the comparison process as they relate to multiple types of comparison. The relevance of our multi-standard approach was supported by the finding that aversive and appetitive comparisons were significantly related to appearance evaluations and were significant

predictions of physical self-concept, body satisfaction, self-esteem, psychological well-being, and depression over and beyond physical appearance social comparison and body satisfaction. These results are in line with similar recent findings on the role of multi-standard comparative thinking in well-being (Morina et al., 2022).

Strength and Limitations

The CSS-A represents the first measure in context of perceived appearance to assess multiple types of comparison and the engendered affective impact. The results are based on a large sample and suggest that comparison frequency, discrepancy, and affective impact should best be assessed separately for appetitive and aversive comparisons. Yet, we also note some limitations. First, convergent validity regarding standards was limited given the lack of comparable measures that measure multiple comparison types. Second, we did not assess test–retest reliability and hence it remains for future studies to examine its sensitivity to change. Another potential limitation is that the CSS-A focussed on upward and downward comparisons only, omitting lateral comparisons. Yet, while lateral social comparisons are frequently applied (Gerber et al., 2018), we suspected lateral comparisons conducted in daily life to be less reliably recalled, given that they are mostly associated with lower levels of engendered emotions. Finally, reporting comparisons that had occurred during the last 3 weeks may have been limited by lack of awareness of non-salient comparisons, lack of recall of relevant comparisons, or denial of aversive comparisons.

Research Implications

The study findings support the notion that multiple comparison types share conceptual parallels and play an important role in appearance-related evaluations and the engendered emotions. Overall, the two-factor model suggests that the CSS-A is best defined as consisting of two independent subscales, i.e., aversive and appetitive. A such, these to subscales may be used independently, depending on the research question. In addition, both subscales distinguish between frequency, discrepancy, and affect. Note that frequency is handled as the core component of comparative behaviour to the effect that discrepancy and affect presuppose the initiation of comparative behaviour (which the CSS-A defines as comparison frequency). To better understand comparison as a process, all three comparison components need to be assessed. However, in some cases researchers might be interested in frequency alone or in combination with discrepancy or affect only, depending on the research question.

While the CSS-A assesses appearance comparisons, a similar approach can be easily applied to assess other relevant facets of self-perception, such as well-being (Morina

et al., 2022). The CSS can also be adjusted to measure more specific types of comparison, such as trauma-related counterfactual comparisons. Additionally, the CSS may also be used with experience sampling methods, which would provide more accurate within-person data. Importantly, future research needs to examine the validity of the CSS-A in individuals with eating disorders and body dysmorphic disorder. The CSS-A may then also be used to examine potential differences between patients with these conditions and healthy participants with respect to the frequency of comparison types (e.g., more or less use of social comparisons relative to temporal comparisons), comparison direction (upward comparisons relative to downward comparisons), comparison discrepancy, or the engendered reactions. Moreover, future studies should investigate the role of ethnicity or membership in certain cultural groups in appearance-related comparisons. Finally, experimental studies are needed to investigate the differential impact of different comparison types on appearance and engendered reactions. Future research needs to consider, however, that the frequency of the different types of comparison and affective impact will likely differ depending on the comparison dimension as well as contextual and personal conditions (Morina, 2021).

Conclusion

In sum, the current approach to defining and assessing relevant components of multi-standard comparisons supports the notion that several types of comparison play a role in appearance judgments and engendered affect. They further indicate relevant differences between aversive and appetitive comparisons. With respect to the CSS-A, the present investigation provides preliminary evidence for its reliability and validity. Continued evaluation of the scale in diverse samples and designs should prove beneficial.

Declarations

References

Albert, S. (1977). Temporal comparison theory. *Psychological Review, 84*(6), 485–503. https://doi.org/10.1037/0033-295X.84.6.485

Allan, S., & Gilbert, P. (1995). A social comparison scale: Psychometric properties and relationship to psychopathology. *Personality and Individual Differences, 19*(3), 293–299.

Anson, M., Veale, D., & Miles, S. (2015). Appearance comparison in individuals with body dysmorphic disorder and controls. *Body Image, 15*, 132–140.

Asparouhov, T., & Muthén, B. (2010). Weighted least squares estimation with missing data. Retrieved January 28, 2022, from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.310.692&rep=rep1&type=pdf

Barnes, M., Abhyankar, P., Dimova, E., & Best, C. (2020). Associations between body dissatisfaction and self-reported anxiety and depression in otherwise healthy men: A systematic review and meta-analysis. *PLoS ONE, 15*(2), e0229268.

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*, 149. https://doi.org/10.3389/fpubh.2018.00149

Broomhall, A. G., Phillips, W. J., Hine, D. W., & Loi, N. M. (2017). Upward counterfactual thinking and depression: A meta-analysis. *Clinical Psychology Review, 55*, 56–73.

Brown, T. A., Cash, T. F., & Mikulka, P. J. (1990). Attitudinal body-image assessment: Factor analysis of the body-self relations questionnaire. *Journal of Personality Assessment, 55*(1–2), 135–144.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*(2), 230–258.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464–504.

Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis*. Psychology Press.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117–140. https://doi.org/10.1177/001872675400700202

Fleming, J. S., & Courtney, B. E. (1984). The dimensionality of self-esteem: II. Hierarchical facet model for revised measurement scales. *Journal of Personality and Social Psychology, 46*(2), 404.

Gerber, J. P., Wheeler, L., & Suls, J. (2018). A social comparison theory meta-analysis 60+ years on. *Psychological Bulletin, 144*(2), 177–197. https://doi.org/10.1037/bul0000127

Grynberg, D., Nandrino, J.-L., Vermeulen, N., Luminet, O., Duclos, J., Goffinet, S., & de Timary, P. (2020). Schadenfreude, malicious and benign envy are associated with low body mass index in restrictive-type anorexia nervosa. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity, 25*(4), 1071–1078.

He, J., Sun, S., Zickgraf, H. F., Lin, Z., & Fan, X. (2020). Meta-analysis of gender differences in body appreciation. *Body Image, 33*, 90–100.

Helm, F., Abele, A. E., Mueller-Kalthoff, H., & Möller, J. (2017). Applying dimensional comparison theory to the fundamental dimensions of social judgment-agency and communion. *Learning and Individual Differences, 54*, 116–125.

Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. Tory Higgins & A. W. Kruglanski (Eds.), *Social psychology* (pp. 133–168). Guilford Press.

Hill, E. M., & Nolan, M. T. (2021). Examining eating disorder-related social comparison orientation and body dissatisfaction in the relationship between fit ideal internalization and drunkorexia engagement. *Eating Behaviors, 41*, 101480.

Hoppen, T. H., Heinz-Fischer, I., & Morina, N. (2020). If only… A systematic review and meta-analysis of counterfactual, social and temporal comparisons in PTSD. *European Journal of Psychotraumatology, 11*(1), 1737453. https://doi.org/10.1080/20008198.2020.1737453

Hoppen, T. H., & Morina, N. (2021). I'd do much better, if only … counterfactual comparisons related to traumatic life events. *Cognition & Emotion, 35*, 409–416. https://doi.org/10.1080/02699931.2020.1832967

Horndasch, S., Heinrich, H., Kratz, O., Mai, S., Graap, H., & Moll, G. H. (2015). Perception and evaluation of women's bodies in adolescents and adults with anorexia nervosa. *European Archives of Psychiatry and Clinical Neuroscience, 265*(8), 677–687.

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

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*(2), 136–153.

Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B. W., Berry, J. T., & Mokdad, A. H. (2009). The PHQ-8 as a measure of current depression in the general population. *Journal of Affective Disorders, 114*(1–3), 163–173.

Lewin, K. (1951). *Field theory in social science: Selected theoretical papers*. Harpers.

McCarthy, P., Meyer, T., Back, M., & Morina, N. (submitted). How we compare: appearance-related comparison standardsassociated with wellbeing.

McCarthy, P., & Morina, N. (2020). Exploring the association of social comparison with depression and anxiety: A systematic review and meta-analysis. *Clinical Psychology & Psychotherapy, 27*, 640–671. https://doi.org/10.1002/cpp.2452

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525–543.

Millsap, R. E. (2012). *Statistical approaches to measurement invariance*. Routledge.

Möller, J., & Husemann, N. (2006). Internal comparisons in everyday life. *Journal of Educational Psychology, 98*(2), 342.

Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review, 120*(3), 544–560. https://doi.org/10.1037/a0032459

Morina, N. (2021). Comparisons inform me who I am: A general comparative processing model of self-perception. *Perspectives on Psychological Science, 16*(6), 1281–1299. https://doi.org/10.1177/1745691620966788

Morina, N., Meyer, T., & Sickinghe, M. (2022). How do I know how I am doing? Use of different types of comparison in judgment of well-being in patients seeking psychological treatment and healthy controls. *Applied Psychology: Health and Well-Being*. https://doi.org/10.1111/aphw.12339

Myers, T. A., & Crowther, J. H. (2009). Social comparison as a predictor of body dissatisfaction: A meta-analytic review. *Journal of Abnormal Psychology, 118*(4), 683.

O'Brien, K. S., Caputi, P., Minto, R., Peoples, G., Hooper, C., Kell, S., & Sawley, E. (2009). Upward and downward physical appearance comparisons: Development of scales and examination of predictive qualities. *Body Image, 6*(3), 201–206.

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance, 17*, 22–27.

Quittkat, H. L., Hartmann, A. S., Düsing, R., Buhlmann, U., & Vocks, S. (2019). Body dissatisfaction, importance of appearance, and body appreciation in men and women over the lifespan. *Frontiers in Psychiatry, 10*, 864.

R Core Team (2021). *R: A language and environment for statistical computing v. 4.0.1*. Vienna, Austria. Retrieved January 28, 2022, from https://www.R-project.org/

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

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5-12 (BETA). *Journal of Statistical Software, 48*(2), 1–36.

Rye, M. S., Cahoon, M. B., Ali, R. S., & Daftary, T. (2008). Development and validation of the counterfactual thinking for negative events scale. *Journal of Personality Assessment, 90*(3), 261–269.

Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology, 69*(4), 719.

Schaefer, L. M., & Thompson, J. K. (2018). The development and validation of the physical appearance comparison scale-3 (PACS-3). *Psychological Assessment, 30*(10), 1330–1341. https://doi.org/10.1037/pas0000576

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology, 53*(1), 1–26. https://doi.org/10.1016/j.cogpsych.2005.10.003

Stice, E., Hayward, C., Cameron, R. P., Killen, J. D., & Taylor, C. B. (2000). Body-image and eating disturbances predict onset of depression among female adolescents: A longitudinal study. *Journal of Abnormal Psychology, 109*(3), 438.

Tantleff-Dunn, S., Barnes, R. D., & Larose, J. G. (2011). It's not just a "woman thing:" The current state of normative discontent. *Eating Disorders, 19*(5), 392–402.

Thompson, J. K., Heinberg, L. J., Altabe, M., & Tantleff-Dunn, S. (1999). *Exacting beauty: Theory, assessment, and treatment of body image disturbance*. American Psychological Association.

Thompson, J. K., Heinberg, L., & Tantleff-Dunn, S. (1991). The physical appearance comparison scale. *The Behavior Therapist, 14*, 174.

Tversky, A. (1972). Elimination by aspects. *Psychological Review, 79*(4), 281–299.

Vlaev, I., Chater, N., Stewart, N., & Brown, G. D. A. (2011). Does the brain calculate value? *Trends in Cognitive Sciences, 15*(11), 546–554. https://doi.org/10.1016/j.tics.2011.09.008

Vocks, S., Busch, M., Grönemeyer, D., Schulte, D., Herpertz, S., & Suchan, B. (2010). Neural correlates of viewing photographs of one's own body and another woman's body in anorexia and bulimia nervosa: An fMRI study. *Journal of Psychiatry and Neuroscience, 35*(3), 163–176.

Wilson, A. E., & Shanahan, E. (2020). Temporal comparisons in a social world. In J. Suls, R. L. Collins, & L. Wheeler (Eds.), *Social comparison, judgment, and behavior* (pp. 309–344). Oxford University Press.

Wood, J. V. (1996). What is social comparison and how should we study it? *Personality and Social Psychology Bulletin, 22*, 520–537.