Psychological Bulletin

A Social Comparison Theory Meta-Analysis 60+ Years On

J. P. Gerber, Ladd Wheeler, and Jerry Suls

Online First Publication, November 16, 2017. http://dx.doi.org/10.1037/bul0000127

Psychological Bulletin

0033-2909/17/$12.00 http://dx.doi.org/10.1037/bul0000127

A Social Comparison Theory Meta-Analysis 60+ Years On

J. P. Gerber
Gordon College

Ladd Wheeler
Macquarie University

Jerry Suls
National Cancer Institute, Bethesda, Maryland

These meta-analyses of 60+ years of social comparison research focused on 2 issues: the choice of a comparison target (selection) and the effects of comparisons on self-evaluations, affect, and so forth (reaction). Selection studies offering 2 options (up or down) showed a strong preference (and no evidence of publication bias) for upward choices when there was no threat; there was no evidence for downward comparison as a dominant choice even when threatened. Selections became less differentiable when a lateral choice was also provided. For reaction studies, contrast was, by far, the dominant response to social comparison, with ability estimates most strongly affected. Moderator analyses, tests and adjustments for publication bias showed that contrast is stronger when the comparison involves varying participants' standing for ability (effect estimates, −0.75 to −0.65) and affect (−0.83 to −0.65). Novel personal attributes were subject to strong contrast for ability (−0.5 to −0.6) and affect (−0.6 to −0.7). Dissimilarity priming was associated with contrast (−0.44 to −0.27; no publication bias), consistent with Mussweiler (2003). Similarity priming provided modest support for Collins (1996) and Mussweiler (2003), with very weak assimilation effects, depending on the publication bias estimator. Studies including control groups indicated effects in response to upward and downward targets were comparable in size and contrastive. Limitations of the literature (e.g., small number of studies including no-comparison control conditions), unresolved issues, and why people choose to compare upward when the most likely result is self-deflating contrast are discussed.

Public Significance Statement
This article summarizes 60+ years of social comparison research and shows that people generally choose to compare with people who are superior to them in some way, even in the presence of threat to self-esteem, and that these comparisons tend to result in worsened mood and lower ability appraisal. Comparisons with proximal persons and on novel dimensions heighten these effects.

Keywords: assimilation, contrast, social comparison, threat, self-evaluation, meta-analysis

Social comparison (Festinger, 1954) can be defined as the "process of thinking about information about one or more other people in relation to the self" (Wood, 1996). The phrase "in relation to the self" means that the comparer looks for or notices similarities or differences from the target of comparison on some dimension. The dimension can be anything on which the comparer can notice similarity/difference. We would usually expect the comparer to react in some way to the existence of the similarity/difference with a change in self-evaluation, affect, or behavior.

Suppose that a young woman discovers that she has an emotional intelligence (EI) score of 10 on a 15-point scale that she found on the Internet. One of her good friends seems really mature and emotionally intelligent, and a second, although vivacious and good fun to be with, seems a bit immature. Whose EI score would she most like to compare her own score with? Comparing with her immature friend would probably boost her ego by making her look superior, but comparing with her mature friend might also boost her ego by showing that she scored as high or almost as high as the mature friend. The question becomes even more interesting if we consider the context. If our young woman's self-esteem has been recently threatened by some remark or event, would this threat push her comparison in an upward (mature friend) or downward (immature friend) direction? The comparison selection made by the young woman is the kind of question we will examine in the first of the meta-analyses presented in this article.

J. P. Gerber, Department of Psychology, Gordon College; Ladd Wheeler, Department of Psychology, Macquarie University; Jerry Suls, Office of the Associate Director, Behavioral Research Program, National Cancer Institute, Bethesda, Maryland.

This work was supported in part by an Australia Research Council Discovery Grant and a Gordon College Initiative Grant. We thank Mike Borenstein, Wolfgang Vichtbauer, Mike Jones, John Ingemi, Michelle Lee, Mason Ostrowski, Maria Street, Isabel Taussig, and Liz Tipton for their help and counsel with various aspects of this study. We are also indebted to several study authors who made available supplementary data.

The views and opinions expressed in this article are those of the authors and do not necessarily represent the views of the National Institutes of Health or any other governmental agency.

Correspondence concerning this article should be addressed to J. P. Gerber, Department of Psychology, Gordon College, Wenham, MA 01984. E-mail: jp.gerber@gordon.edu

The second meta-analysis we will present deals with reactions to social comparison. In these studies, we would not ask or allow the young woman to select a target to compare herself to but rather would present her with social comparison information about a target or targets. In such studies, the investigator controls the identity and characteristics of the target in order to better understand the effects of social comparison. We will look at whether social comparisons are more likely to lead to assimilation or contrast, and what variables moderate assimilation/contrast. Figure 1 illustrates what we mean by contrast and assimilation. The baseline of the self-evaluation is obtained from a premeasure or from a control condition in which the participant is not exposed to either an upward or downward comparison. An upward comparison is one in which the comparison standard is better off than the comparer, and a downward comparison is with a standard who is worse off. In the two instances of assimilation, the comparer's self-evaluation moves toward the comparison standard, becoming more positive after an upward comparison and more negative after a downward comparison. Conversely, in the two instances of contrast, self-evaluation moves away from the standard, becoming more negative after an upward comparison and more positive after a downward comparison.

Some Relevant Theoretical Perspectives

Downward Comparison Theory

Wills (1981) proposed in his theory of downward comparison that threat produces downward comparison in an attempt to restore self-esteem. Wills argued that comparisons are generally upward (a conclusion based largely on the selection studies to be reviewed in the first meta-analysis) but that when self-esteem is threatened, comparison changes to a downward direction (Hakmiller, 1966) to restore self-esteem. Wills also predicted that people with low self-esteem would be particularly prone to downward comparison when self-esteem is threatened. Wills' theory is actually a two-part theory, one part predicting downward selection (when threatened) and the second part predicting a positive reaction after downward comparison. Other theories (e.g., Wood, 1989) also assert that downward comparisons can boost self-evaluations so it is the first part that distinguishes Wills' downward comparison theory. In the first meta-analysis presented in this article, we will examine studies of comparison selection and whether threat has the proposed effect of producing downward comparison.

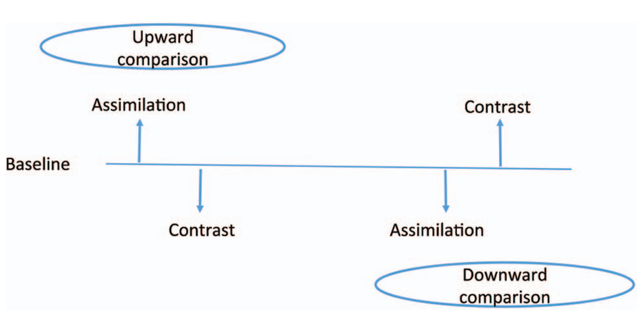

Figure 1. Assimilation and contrast following upward and downward comparison. See the online article for the color version of this figure.

Construal Theory and Upward Social Comparisons

Collins (1996, 2000) proposed that people may assimilate their self-evaluation upward to those who are better off. Collins was influenced by the early work on comparison selection using the rank-order paradigm (Thornton & Arrowood, 1966; Wheeler, 1966) showing that people generally choose to compare upward. Collins argued that an upward comparison can be construed as showing similarity to the better-off target, leading the person to elevate their own self-worth to be in the same category as the target. This is most likely to happen if the person making the comparison initially expects to be similar to the target and is thus primed to construe any difference as slight or nonexistent. Downward assimilation is unlikely to occur because people do not expect or want to be similar to worse-off others and will not construe differences to be slight. Empirically, high self-esteem and shared distinctiveness with the target (Brewer & Weber, 1994) are the two factors most clearly linked to upward assimilation, according to Collins.

The Selective Accessibility Model (SAM)

Mussweiler (2003; Mussweiler & Strack, 2000) has also proposed a model emphasizing assimilation, the selective accessibility model (SAM). When faced with a possible social comparison, a person makes a tentative and rapid judgment of similarity or dissimilarity to the comparison target. This can be based on any information we have about the target, although Mussweiler argues that the default hypothesis is that of similarity. Then there is a cognitive search for information consistent with the preliminary hypothesis of similarity or dissimilarity (e.g., Klayman & Ha, 1987). In the case of self-other comparisons, whether one searches for similarity information or dissimilarity information, it should be easy to find information that is consistent because self-concepts are remarkably rich and complicated. That information then becomes selectively accessible when we make judgments about ourselves. If we have searched for information that we are similar to the standard, we are likely to assimilate our self-evaluations toward the target. If we have searched for information that we are dissimilar to the target, we are likely to contrast our self-evaluations away from the target.

Social comparison, according to SAM, does not just increase the accessibility of standard consistent knowledge but also provides a reference point against which this knowledge is evaluated (Upshaw, 1978; Manis, Biernat, & Nelson, 1991). Comparing our scientific success to that of an American Psychological Association Distinguished Scientist may make knowledge of our own scientific achievements more accessible, which should lead to assimilation, but it also provides a reference point (the Distinguished Scientist) against which to judge our success, which should lead to contrast. Absolute or objective questions such as "How many publications do you have in top journals?" should show the effects of assimilation (due to greater accessibility of our scientific achievements), whereas subjective questions such as "How distinguished is your scientific career?" should lead to both contrast (due to the high reference point) and to assimilation (due to greater accessibility of our scientific achievement) . Thus, the predictions of SAM can most clearly be tested with objective questions (Mussweiler, 2003).

Target Immediacy

Local information tends to be more highly weighted in self-evaluation than distant information (Zell & Alicke, 2010). For example, although comparison information about other students at a university may have effects on a person's self-evaluation, those effects can easily be removed or muted by comparison information about the other students in one's dormitory, and those effects can be removed or muted by comparison information about the students in one's suite. A well-known example from education is the frog-pond effect (Davis, 1966; Huguet et al., 2009; Marsh & Parker, 1984), which shows that academic self-concept is lower in highly selective high-schools than in less selective high-schools (after controlling for student academic ability). Students in elite high-schools compare themselves to their fellow students rather than to the outside world and suffer in self-concept as a result. Zell and Alicke (2010) believe this to be true because humans evolved in small groups and are habitually exposed to peer comparisons during development. We coded how close the comparison target was to the comparer, reasoning that comparison with a local target should have larger effects than comparison with a distant target.

Self-Evaluation Maintenance Model (SEM)

One theoretical framework, Tesser's (1988) SEM, is not tested in this article so an explanation for this apparent omission is appropriate. SEM predicts how people maintain self-evaluation under the confluence of three interacting independent variables: performance of the self relative to another, psychological closeness to the other, and the relevance of the comparison dimension to the individual's self-definition.

Under the SEM, comparison occurs when the actor is outperformed by a close other on a relevant dimension. In this case, upward contrast will result, and the actor's ability-evaluation will be diminished. The comparison process always leads to downward contrast from a superior other. This prediction of upward contrast is not unique to SEM and is tested as a matter of course in our analysis.

If, however, the dimension is not relevant to the actor's self-definition, the actor can bask in the reflected glory of her close friend (i.e., what Tesser called "reflection"). Her self-esteem and affect may increase (but her self-evaluation on the performance dimension will not increase). Unlike Tesser, some writers unfortunately have equated reflection with assimilation. Assimilation occurs when one's drive for distinction leads to a perceived similarity in ability with a high-performing other. Reflection occurs when one gives up any claim to distinction on the comparison dimension and basks in the connection (or in Heider's [1958] terms, "unit relationship") with a psychologically/emotionally close other. Thus, the reflection prediction of the SEM does not rely on social comparison per se.

Where possible, however, we included any SEM-derived studies that met our inclusion criteria (e.g., Blanton, Crocker, & Miller, 2000). But Tesser's model does not use self-evaluation as a dependent variable and instead uses other dependent variables such as perceptions of the other, changes in relevance of a self-definition, and closeness to the other. As such, there are few SEM studies that directly involve social comparison, echoing Pleban and Tesser's (1981) suggestion that the ". . . comparison process is not the same as that discussed by Festinger (1954)" (p. 279). We were unable to code the studies for Tesser's variable of relevance (because it was rarely a manipulated variable), but we did code them for immediacy of the comparison target, an approximation of Tesser's variable of closeness.

Method Used in Studying Social Comparison

Selection Method

One way to study social comparison is to look at the targets selected for comparison. One of the earliest and most commonly used methods has come to be known as the *rank order paradigm*. In the original study (Wheeler, 1966) seven male participants were tested together to determine which students would take a special small psychology seminar in place of the standard lecture course the following quarter. Each participant was led to believe that he was at the middle rank (4) in the group. He was told his score and the approximate scores of the highest and lowest ranks. He was then allowed to select one person in the group whose score he would most like to know. It is clear that the participant could make an upward or a downward comparison and could choose the most similar other (Ranks 3 or 5) or a more distant other (Ranks 2 or 6). Participants compared strongly upward and compared predominantly with the most similar upward target. A correlation showed that those who thought they were closer in score to the person above them in the rank order than to those below them in the rank order were more likely to select an upward target—leading to the conclusion that those who compared upward were attempting to confirm their assumed similarity upward.

The rank order paradigm has been used with various modifications. Hakmiller (1966), for example, predicted a certain rank order from the Minnesota Multiphasic Personality Inventory (MMPI) but then gave the participant a much higher score than anticipated when the trait was presumably measured with the more reliable "thematic apperception slides and galvanic skin measurement." The trait was hostility toward one's parents, described as very negative (high threat) or as not very negative (low threat). Then they were given their quite high score from the more reliable measurement. Participants were allowed to see the score of one other person. High threat participants more than Low Threat chose to see the score of the person predicted to have the highest score in the group (most hostile). This experiment was the first to show that threat can lead to comparison with someone worse off than oneself.

Other laboratory selection studies that do not use the rank order paradigm nevertheless use methods with similar features. Almost all selection studies provide information that some targets are superior or should be superior to the participant, while other targets are inferior. Some studies also allow choice of targets who are neither upward nor downward but rather lateral (or the same) to the participant. Diary studies such as Wheeler and Miyake (1992) also provide usable data on comparison selection.

Reaction Method

In this group of studies, participants are exposed to comparison targets, and participants' reactions are the dependent variable. One of the earliest and best-known of these studies is the Mr. Clean/Mr. Dirty experiment (Morse & Gergen, 1970). Participants who were filling out job applications found themselves sharing a table with either Mr. Clean, a well-dressed and highly organized college student or Mr. Dirty, a dishevelled and disorganized young man. The change in self-esteem from a pretest was positive for those exposed to Mr. Dirty and negative for those exposed to Mr. Clean. In other words, participants contrasted themselves against their fellow job applicant.

Another example of a reaction study is Mussweiler (2001). Similarity/difference was procedurally primed (Smith & Branscombe, 1987) by having the participants examine two drawings of a person in a room containing various objects. Participants were asked to find as many similarities (or differences) as possible between the two drawings. They then read about another college student who was adjusting very well or poorly to college. The dependent variable consisted of two objective questions that assessed self-evaluations of their adjustment to college: how often they went out each month and how many friends they had at college. Participants primed with similarity rated their adjustment higher when exposed to the successful student, but those primed with difference rated their adjustment higher when exposed to the poorly adjusting student.

Narrative Method

A third category of studies is the narrative approach (Wood, 1996), having the three major subsets of (1) global self-reported comparisons, (2) self-recorded social comparison diaries, and (3) free response methods.

1. Global self-report. Participants are asked who they normally compare with or have compared with in the past. This method is little used because of memory and bias concerns.

2. Self-recorded social comparison diaries. Participants can be asked to complete a diary each day about their social comparisons and other relevant events. This is called an interval-contingent self-recording (Wheeler & Reis, 1991). One can also use what is called event-contingent self-recording, in which the occurrence of an event (such as a social comparison) triggers the self-recording. Both interval and event contingent diaries were included in the analyses reported in this article. Using the Rochester Social Comparison Record, Wheeler and Miyake (1992) had participants keep an ongoing record of every social comparison they made over 2 weeks. Participants indicated the dimension of comparison (appearance, academic, etc.), their mood before and after the comparison, their relationship to the target person (friend, acquaintance, etc.), and whether the participant was superior, lateral, or inferior to the comparison target. Thus participants were selecting comparison targets from their everyday life rather than from the possibilities given by an experimenter. Under these conditions, 43.2% of the comparisons were downward, 30.4% were upward, and 26.4% were lateral. Downward comparisons were made when respondents were in relatively positive mood; upward comparisons, when respondents were in a relatively negative mood. These results are directly opposite to those predicted by downward comparison theory (Wills, 1981) and instead support a selective affect-cognition priming model (Bower, 1981, 1991). In this model, affect primes cognitions, such that dysphoria primes cognitions of worthlessness. The selective part of the model is that the cognitions of worthlessness are applied primarily to the self. A dysphoric person is likely to see other as superior and to make upward comparisons.

3. Free response methods. An early use of free response measures is the breast cancer research done by Wood, Taylor, and Lichtman (1985), in which a social comparison was recorded whenever a woman made a comparative statement about how she was doing with respect to her illness. Women who had been treated surgically for their breast cancer (median time since surgery was 25.5 months) were interviewed face-to-face for about two hours about all aspects of their cancer experience. The women were most likely to make spontaneous statements indicating that others were coping less well than they were. "I have never been like some of those people who have cancer and they feel well, this is it, they can't do anything, they can't go anywhere. . . . I just kept right on going" (Wood et al., 1985, p. 1173). Downward conclusions like this were more frequent than upward conclusions on adjustment, physical status, and life situations.

   There was no direct measure of comparison, so the data are really just the women's conclusions about their adjustment (see Tennen, McKee, & Affleck, 2000, pp. 470–472). As this study measured coping, and did not directly measure social comparison, it does not meet our inclusion criteria presented in the following text. Most subsequent studies, although inspired by Wood et al., had respondents report or make actual comparison selections (these studies are included in the review).

We have attempted in this brief introduction to mention the major issues investigated by our meta-analysis. Other matters dealt with will be discussed in context. The meta-analysis has two main sections: (1) The first section deals with selection studies, in which the major question concerns the influence of threat on direction of comparison. We also looked at selection of the most similar comparison targets versus more dissimilar targets and at differences between laboratory and field studies. (2) The second section deals with reaction studies, in which we concentrate on self-evaluative changes as a result of upward and downward comparison. In this latter section, we focus on moderators of the self-evaluative changes, hoping to find those moderators that lead to assimilation versus contrast.

Meta-Analysis of Selection Studies

Method

Database construction. For all sections of this report, we utilized one main search, identifying relevant articles in four ways: a database search using PsycINFO, Wheeler's personal database, the SPSP discussion list, and searching the curriculum vitae of known social comparison researchers. Notes on each method are presented in the following text, and full PRISMA diagram is presented in Figure 2.[1,2]

1. PsycINFO database search. We searched "Social Comparison" in the title and abstract, and also as a keyword in PsycINFO on December 29, 2008, then again on October 12, 2013 (to cover the period since December 29, 2008). After this, we switched to an EBSCOHost alert for the same search. The 2008 search generated 2,848 abstracts, whereas the 2013 search led to an additional 290. The abstracts of these were read and all articles that might possibly be about social comparison were obtained.

2. Wheeler's collection. Wheeler has maintained filing cabinets of major social comparison articles and unpublished data sent by other researchers. These were included in the main database and also used for our initial coding run.

3. The SPSP discussion list was emailed in 2008 for additional unpublished data.

4. Many social comparison researchers post their publications online, so we used these online lists to double-check for any additional articles.

Database completeness. Missing data were followed up via email requests to researchers. Across all sections of this article, this approach resulted in additional data for 35 articles. Any article with incomplete data was included as far as possible.

Inclusion criteria for selection studies. In order to make comparisons, people must have a target with which to compare. These targets are either given to people or they need to be selected by the individual. Studies in which participants choose a target are called selection studies. Two comparison choice variables that selection studies often measure are direction and distance. Direction refers to whether the target is perceived as better than, worse than, or equal to the comparer, usually with respect to the dimension in question. These are referred to as upward, downward, and lateral targets (or comparisons), respectively. Distance refers to how close the target is to the comparer on the comparison dimension. The target may be only slightly better or worse than the comparer (a near comparison) or greatly better or worse (a far comparison).[2]

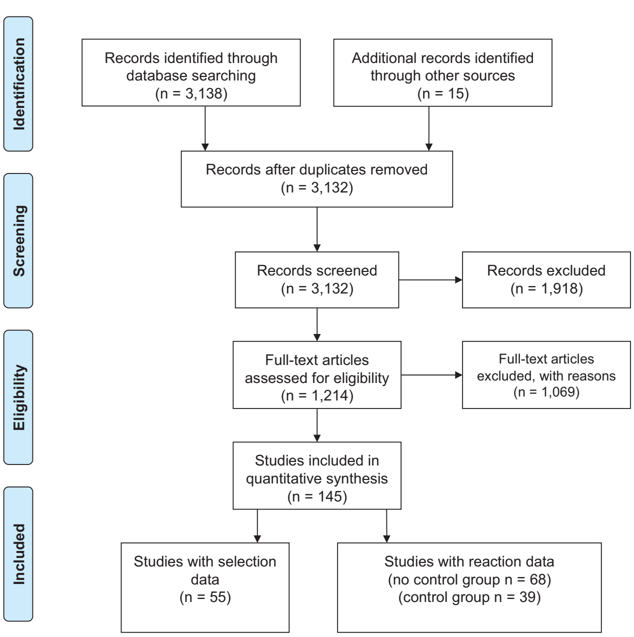

Figure 2. PRISMA diagram. See the online article for the color version of this figure.

In this analysis are studies that reported choosing a single target. These studies could be either field studies or lab experiments. In field studies, participants completed either a diary of comparisons, or a questionnaire that assessed selection of targets.[3] We found 55 articles that met the previously mentioned criteria, with a total of 101 effect sizes (see Appendix B in the online supplemental material).

Selection Study Coding

Comparison direction. Raw counts of up, down, and lateral comparisons were recorded. These were then recoded into proportions for the tables presented in the following text.

Setting of study. As is common for all areas of psychological research, the setting of a selection study may influence the targets participants choose. Laboratory studies may encourage a tendency toward upward comparison that is not present in field studies because expressing a desire to learn about superior performers may be considered more socially desirable and therefore more likely when face-to-face with an experimenter. Therefore, we tested the difference in selection targets between lab and field studies. Field studies were further divided into event-contingent and interval-contingent diary studies.

Threat. The presence of threat might influence selection choices. As outlined in our introduction, downward comparison theory suggests a flight to downward choices when threat is present. There are multiple routes to a threatening comparison. Threat was coded in three categories: no threat, threat induced by the experimenter, and health-related comparisons (for those with serious illnesses such as cancer). As per Hakmiller's (1966) method, experimentally induced threat required giving participants negative feedback about their own abilities or other attributes before asking them to choose a comparison target.

Analytic approach. We split the analysis into two groups: studies with a lateral choice option, and studies that only offered upward and downward options. Although many studies yielded more than one effect size, the only interdependence between effect sizes was from being in the same article and/or the same lab. As such, we used a multilevel model to analyze our data, with each article having its own random effect.

Our choice of analytic model was limited to mixed-effects binomial logistic regression because several observed proportions were close to, or exactly, one (i.e., nearly everyone chose upward). The logit model, a typical transformation for proportional data,

[1] See supplemental online materials (https://osf.io/k27hm/) for the full list of articles.

[2] Distance is not analyzed in this article because of difficulty agreeing on a satisfactory coding rule.

[3] There was also one study (Buunk, Schaufeli, & Ybema, 1994) that had both diaries and questionnaires within one study. This was removed from the analyses examining the effects of setting on comparison choices.

was not considered appropriate here because the logit is skewed at values close to 1. The models were run using using the GLMM4 command from the *metafor* library for R.

The necessary choice of logistic regression reduced the options for publication bias corrections. We examined funnel plots and the Egger's regression test only. Other forms of publication bias correction are not available within the logistic regression context (e.g., Comprehensive Meta-Analysis only handles proportions as a logit model, *metafor* has publication bias corrections only for the logit model). Beyond these immediate difficulties, funnel plots for proportional data are best done with weighting by sample size, not standard error (Hunter et al., 2014). These variations are not directly implemented in R so we used syntax to create the Egger's tests.

Reporting conventions. We use k to refer to the number of effect sizes in each analysis. The number of independent studies contributing to each effect size is listed at the bottom of each table.

Results

Studies offering only upward versus downward options. There was an overall preference for upward comparison ($Z = 6.19$, $p < .001$). Upward choices were made 75.6% of the time, while downward choices were made 24.4% of the time. To compute the two-way interaction between setting and threat, the lab/no threat was used as a reference category and logistic models were computed (see Table 1). This model was significant, $Q_m(4) = 20.75$, $p < .001$. Threat tended to reduce upward choices in the lab slightly, but they were still the dominant choice. Moving to the field (event-contingent or interval-contingent diaries) settings led to a reduction in upward comparisons ($Z_{\text{interval-contingent-diary}} = -2.68$, $p = .007$, $Z_{\text{event-contingent-diary}} = -4.09$, $p < .001$), but the downward option was not preferred. For comparison selections in the field under no-threat, there was no preference for upward or downward comparisons. Even with threat, downward comparison was not a dominant preference in field studies. Upward comparisons were preferred. There was no evidence of bias: regression tests for publication bias were nonsignificant, ($Q_{\text{lab-induced-threat}}(1) = 1.43$, $p = .232$, $Q_{\text{lab-no-threat}}(1) = 0.92$, $p = .336$, $Q_{\text{diary-studies}}(1) = 1.66$, $p = .198$). Publication bias for the field recall studies was not calculable due to low k. The cell averages are reported in Table 1.

Studies with upward, downward, and lateral options. Studies offering the three options were analyzed as a series of binomial logistic mixed-effects meta-analytic models (i.e., upward vs. downward, upward vs. lateral, downward vs. lateral). Overall, approximately 46% of the selections were upward, 18.5% were lateral, and 35.6% were downward (see Table 2). The difference between upward and lateral choices was marginally significant ($Z = 1.89$, $p = .058$), Wald's $\chi^2(44) = 2393.02$. The difference between upward and downward was significant ($Z = 2.48$, $p = .013$), Wald's $\chi^2(44) = 440.80$; the proportions of downward and lateral choices did not differ significantly ($Z = -.18$, $p = .859$), Wald's $\chi^2(44) = 2132.15$. The pattern of significance and nonsignificance between the lateral and directional proportions varies with the amount of hetereogeneity in the effect sizes, as shown by the Wald's chi-square values.

Table 1
Associations Between Threat, Study Setting, and Choice of Upward Comparisons

Threat	Lab	Field (interval contingent diary)	Field (event contingent diary)
No threat ($k = 20, 2, 7$)	85%[a]	45%	49%
Health threat ($k = 2$)		82%[a]	
Induced threat ($k = 23$)	74%[a]		

Note. Number of studies for each line are (9, 2, 2), (2) and (10).
[a] Indicates proportion is significantly different from 50%.

There were no significant interactions involving threat or setting for upward versus neutral conditions ($Z = -0.54$, $p = .588$), upward versus downward ($Z = 0.21$, $p = .216$), or for downward versus neutral ($Z = 0.62$, $p = .538$). Publication bias methods for the trichotomous moderators are omitted because they would only attenuate an already nonsignificant result and because the low ks are prohibitive.

Selection results summary. In upward-downward choice studies, upward comparisons were preferred in laboratory settings and were depressed only modestly when threat was present. Field settings were associated with less clearly defined preferences although downward choices never predominated. Thus, there was little support for the idea that threat induces downward comparisons. The addition of a lateral choice to the experimental paradigm reduced the differences between choice preferences although upward selections still had the "edge." Also, when lateral comparison was an option, neither threat or setting moderated selections. One caveat is that the number of studies offering a third option was smaller than for the two-choice paradigm. One speculation is that the lateral choice represents the opportunity to share a "similar-fate" (Schachter, 1959; Wills, 1981; Darley & Aronson, 1966), which competes with the striving to identify with a person of higher standing. If the two (conflicting) motives are salient then the selection process may be complicated for participants.

Meta-Analysis of Reaction Studies Without Control Groups

Method

Inclusion criteria for reaction studies. Reaction studies involve the manipulated exposure to a possible comparison target. In such studies, the participant is presented with comparison information about a target, and the effect of this information on the participant is measured. The precise criteria used for inclusion are listed in the following text.

1. Studies needed to manipulate exposure to a target that was either upward (i.e., better than them) or downward (i.e., worse than them). This criterion excluded opinion comparison studies, which have no clear direction of comparison. The experimental design needed to randomly assign the comparison condition so interview and diary-based studies were excluded. The study had to engage the actual comparison process in the laboratory; recollection of the responses to previous comparisons was not considered appropriate. This criterion excluded studies that relied on any of 14 scale-based measures of social comparison orientation, impact, or frequency (e.g.,the physical appearance comparison scale, Allan &

Table 2
Average Choice When Lateral Choice is Also Available

Direction	Choice
Up (k = 45)	45.9%
Similar (k = 45)	18.5%
Down (k = 45)	35.6%

Note. Number of studies for each line is 33.

Gilbert, 1995; the Iowa-Netherlands Comparison Orientation Scale, Gibbons, & Buunk, 1999) without any direct comparison target. Some body-image social comparison scales have been meta-analyzed by Myers and Crowther (2009).

2. The comparison target in these studies had to be a single, specific person. This criterion excluded studies that included the 'average person' or an 'ideal group member,' and also studies that used multiple targets (e.g., multiple ad ratings, common in body image research) or a group as the target (e.g., boys, girls). This criterion was established to exclude studies about the better-than-average-effect, which is a social projective bias (Alicke, 1985; Krueger & Clement, 1994), and to more accurately gauge the impact of single comparisons. Comparisons with the "average" have different effects to comparisons with individuals (Zell & Alicke, 2010) and have been meta-analyzed by Sedikides, Gaertner, and Vevea (2005).

3. The study had to collect some self-related dependent variable. This might include mood, self-esteem, an ability estimate, performance satisfaction, or some behavioral measure related to the ability in question. This criterion excluded target judgment (i.e., person perception, such as liking) studies. Studies were also excluded if the dependent variable was not directly related to self (e.g., purchase intentions) or was ambiguous in its meaning (e.g., self-reward via tokens). This criterion also excluded some studies on equity and wage fairness.

4. There had to be no intervening manipulations between the presentation of the comparison stimulus and the measurement of the dependent variable.

5. The study had to be a between-subjects design, to assure results were comparable across studies. Within-subject studies, either with prepost measures or multiple conditions, are exceedingly rare (a notable exception is Morse & Gergen, 1970, a between-groups study including repeated measures).

6. The study design had to have at least two groups. The most common laboratory reaction study was one in which an upward and downward target are used without a neutral control, and this type of study is analyzed in the main reaction section. There are also several types of control groups in the literature from no comparison controls, to lateral controls. Studies with control groups are analyzed in a subsequent section on assimilation/contrast.

Excluded studies. As detailed in the preceding text, there are some domains that we did not include. For example, we excluded frequency of comparison, opinion comparison, affiliation studies, and related attribute similarity/dissimilarity. There were several reasons that a study might not meet our inclusion criteria. As outlined in the preceding text, the sources of inadequacy identified included presenting multiple targets, confusing social comparison with other effects such as the better-than-average effect, and using group averages or percentile ranks instead of direct social comparisons.

Diederik Stapel is a Dutch social psychologist who had 58 publications retracted because of suspicion of or admission of faked data. We have of course excluded all retracted studies but have included four studies that have undergone extensive review by Dutch authorities and allowed to stand.[4]

Summary of final reaction database. Sixty-eight articles met the inclusion criteria outlined in the preceding text (see Figure 2 for full PRISMA diagram). The final reaction database consisted of 185 lines of effect sizes for reaction studies (see Appendix C in the online supplemental material).

Coding scheme for dependent variables and moderators. All moderators were derived from reading key theoretical social comparison articles and were decided prior to the main coding and/or the computation of any effect sizes and statistics. The complete set of selected moderators is reported here.

Types of dependent variables. Six types of dependent variables were coded.

1. Ability estimate—where a person gives an estimate of their own standing on the attribute in question; for example, "How smart are you?", "How popular are you?", "How many push-ups can you do in 10 min?"

2. Affect—These were coded from mood scales and items (e.g., PANAS [Watson, Clark, & Tellegen, 1988]; MAACL [Zuckerman, 1960]; happy/sad, angry).

3. Self-esteem—for example, competence, self-concept, self-efficacy. These were mainly from the Heatherton and Polivy (1991) state measure, but a few studies used the Rosenberg (1965) trait measure.

4. Behavior—doing a second task afterward that was related to the comparison (e.g., a trivia quiz after an intelligence comparison, walking down a hall after comparing with an elderly person).

5. Performance satisfaction—People rated how satisfied they were with their performance (e.g., "How satisfied are you with how you did on the test?"). This is not the same as ability; people can easily be satisfied or dissatisfied with a

[4] We did, however, conduct analyses to see what effect Stapel's retracted studies had if they were included. These analyses (available online) revealed that the inclusion of Stapel's work would confirm his prediction that implicit & explicit comparisons have different effects. Without Stapel's data, this effect was only marginal. Beyond this, in no case would Stapel's fraudulent results lead to a new conclusion to the results reported in the following text, although some effects did become nonsignificant and some cell means did show variation in effect size.

performance that reflected high ability. There were relatively few studies that included performance satisfaction so no moderator analysis was feasible.

6. Other. All other dependent variables (DVs) were kept in this category. This category included such variables as global goals and perceived task difficulty. This category was rarely represented in the study sample and remains unanalyzed.

Comparison induction. The method used to induce a comparison varied along two dimensions: explicit/implicit and similarity/difference seeking (in the case of explicit comparisons). An explicit induction involves giving the comparer a target and either asking them to compare or providing an explicit comparison question (as long as this is prior to the dependent variables). Implicit induction involves presentation of a comparison target but the participant is not told to compare with them or given a comparative question. Similarity-seeking occurs when participants are asked to assess similarity to target (or primed to do so). Difference-seeking occurs when participants are asked to assess dissimilarity from target (or primed to do so).

Local/distant target. Local information tends to be more highly weighted in decision making than distant information (Zell & Alicke, 2010). We coded how close the comparison target was to the comparer, reasoning that comparisons with local targets might have larger effects than comparisons with distant targets. Whether the paradigm used a local/in vivo target or a more distant target was coded.

Dimension novelty. We coded for whether the dimension was created by the researcher or was (feasibly or actually) already known to the participant. People should have some anchors for known dimensions, making a comparison less powerful. Unknown dimensions (e.g., where the researcher makes up a new test of some fictitious ability) should have heightened comparison effects.

Varying self versus standard. In order to make participants believe that they are better or worse than others, participants can be given false feedback about their own performance and manipulated information about the target, or given true (or no) feedback about their own performance and manipulated information about the target. We refer to this dichotomy as *varying self* or *varying the standard*, respectively. Varying the self is likely to create uncertainty, thus increasing the desire for social comparison, similar to Festinger's (1954) second hypothesis. For example, if the experimenter changed the feedback for the participant (e.g., said they scored poorly on an intelligence test when they had not), this might lead to uncertainty and a desire for social comparison to evaluate an uncertain attribute, hence heightening the impact of the social comparison. Studies that, instead, vary the standard should have weaker effects.

The use of exemplars. Dijksterhuis et al. (1998) suggested that exposure to exemplars (e.g., Albert Einstein, Claudia Schiffer) leads to contrast whereas category members (e.g., professors, supermodels) lead to assimilation. We coded social comparison studies for whether (1) they used an exemplar (e.g., Einstein, Claudia Schiffer), or (2) they used a nonexemplar (e.g., a smart student, a magazine model). This work has recently become the subject of a multisite replication study because of previous problems with replicability (see https://osf.io/k27hm/).

Objective/subjective dependent variables. Objective or absolute rating scales use objective anchors (e.g., How many pushups can you do?). Subjective or dimensional rating scales use subjective anchors (e.g., How physically fit are you? ranging from "unfit" to "fit"). Mussweiler and Strack (2000) argued that the type of rating scale used might influence the outcome of the comparison when assessing our abilities in relation to a target. Specifically, objective DVs (vs. subjective scales) may be less susceptible to response scale interpretation produced by a comparison target (Biernat et al., 1991) and hence provide stronger evidence for assimilation.

Self–other/other–self direction. Social comparisons that focus on the self first (self-to-other) may lead to more contrast than social comparisons that focus on the target first (other-to-self; Kühnen & Haberstroh, 2004). If the self is the starting point of the comparison and has more elaborated features, which almost by definition, the self does, (Srull & Gaelick, 1983; Tversky, 1977), it should be seen as less similar to the target, hence creating contrast. We coded whether the paradigm made participants focus on the self first, or the target first.

Analytic model. One of the most pressing issues in meta-analysis is to assess the impact of publication bias on meta-analytic findings. In what follows, we describe our approach to both answering substantive theoretical questions and also adequately accounting for the effects of potential publication bias. Our approach was to estimate the effect size in a number of ways and then compute several publication bias correction techniques as a sensitivity analysis to provide a range of plausible estimates for any true underlying effect (see Inzlicht et al., 2015).

Our database required an analysis that included random effects and accounted for multiple endpoints (i.e., multiple dependent measures within one study; Gleser & Olkin, 2009). The possible options were either a multilevel multivariate random-effects model, which required the correlations between dependent variables in each study in order to ensure the variance weights were estimated accurately, or robust variance estimation (RVE) with sandwich estimators (Hedges, Tipton, & Johnson, 2010)[5]. In nearly all cases, the correlations between dependent variables were not available in the published literature. However, we were able to recover 23 out of 32 necessary correlations via email requests to study authors. Given this high success rate, we imputed the remaining 9 correlations with the average of the relevant correlations and ran the multilevel multivariate model. We also ran the RVE model as a comparison, but note that the multivariate analysis is likely to give the most precise measure of the effect size because it utilizes more information than RVE.

Next, we compared these multivariate approaches to a univariate random-effects model. We expected similar effect sizes because fewer than 20% of studies had multiple endpoints. The univariate analysis also gave us a reference point for the publication bias sensitivity analyses (all of which are univariate).

The first analysis was an overall assessment of contrast and assimilation across all dependent variables, and then the testing of the impact of dependent variable type. Our prediction was that ability evaluation, historically the outcome most central to social

[5] There was a sufficient number of studies to conduct RVE, given that the *k* of 68 is larger than the suggested minimum *k* of 40.

comparison theory, would afford the best opportunity to discern reliable effects.

Following the overall assessment and testing of outcome type, we used multilevel metaregression models to examine moderators. We entered the main effects of moderator and variable type, and their interaction, as shown in the following equation:

$$d_{\text{study}} = \alpha + \beta_{\text{DV\_type}} + \beta_{\text{moderatorr}} + \beta_{\text{DV\_type}\,x\,\text{moderator}} + \varepsilon$$

For each moderator, we included random effects at both the study level and dependent variable level. This model allows for simultaneous assessment of the priority of some dependent variable types (in the first coefficient) and whether the moderator had an overall impact (the second coefficient) along with the interaction. Where the interaction was not significant, we dropped it.

To be more precise, our questions were as follows:

1. Does social comparison influence some dependent variables more than others? This was answered in the overall analysis of dependent variable type and also in the first coefficient of each of our regression models.

2. Is there an interaction between dependent variable type and moderator?

3a. (If there is a significant interaction). What are the simple effects of the moderator for each different dependent variable? In other words, what levels of assimilation and contrast exist for different combinations of the moderator and dependent variables?

3b. (If there was no significant interaction). What, if any, is the effect of the moderator? Is assimilation and/or contrast evident at different levels of the moderator?

Results addressing all of these substantive questions were considered in the context of publication bias estimates (described in the following text), to provide a plausible range of effects sizes. As will be seen in the bulk of the tables, the mean ES and 95% CI for the effect size estimates are presented first, followed by the publication-bias corrected estimates and the two inferential tests of publication bias (i.e., Eggers Test and rank-correlation test).

Effect size (ES) meaning (very important, read carefully). The final two questions (3a and 3b) were addressed by computing range estimates of ES; hence clear understanding of our effect size coding is necessary. All effect sizes were coded so that a positive effect size indicates that the experimental condition assigned an upward comparison has more of the attribute than does the group assigned a downward comparison, where Cohen's $d = (M_{up} - M_{down})$/(pooled variance). When examining assimilation and contrast, there are two basic results that would produce the same positive effect size, and two basic results that would produce the same negative effect size. If we use IQ ($M_{POP} = 100$) as an example and ignore the pooled variance, we have the following possibilities for positive effects: ($M_{up} = 115$, $M_{down} = 100$, $d = +15$), indicating upward assimilation, and ($M_{up} = 100$, $M_{down} = 85$, $d = +15$), indicating downward assimilation. Similarly, there are two possibilities for negative effect sizes. ($M_{up} = 85$, $M_{down} = 100$, $d = -15$), indicating upward contrast, and ($M_{up} = 100$, $M_{down} = 115$, $d = -15$), indicating downward contrast. In addition, any given *d* could be a mixture of upward and downward assimilation or upward and downward contrast effects. For example, $M_{up} = 107$, $M_{down} = 92$, $d = .15$, indicating both upward assimilation and downward assimilation. The analyses reported in the next section allow us to determine if assimilation is more or less prevalent than contrast, but it will not allow us to determine if upward assimilation is more or less prevalent than downward assimilation, or if upward contrast is more or less prevalent than downward contrast. There is no way to tease these apart without control groups (we note that the majority of reaction studies lack control groups). In a later section, we will examine those studies that do have control groups.

Other Statistical Issues

Publication bias. A series of publication bias techniques were used as a sensitivity analysis. The Egger's regression test and the rank-correlation tests were used as inferential tests of publication bias; significance in either would be highly suggestive of publication bias.[6] However, both tests are known to be insensitive to some forms of publication bias (Higgins & Green, 2011) so we also ran publication bias estimates in all cases. The publication bias corrected estimates chosen were Vevea and Woods' (2005) weight function selection method, the trim-and-fill method, and the PEESE technique. Of the three, PEESE is likely to be weakest as it often has poor coverage (of the true effect size; Inzlicht, Gervais, & Berkman, 2015; Hilgard, 2015) and performs less well than weight function selection methods (McShane, Böckenholt, & Hansen, 2016). Trim-and-fill may be inaccurate if the data have high heterogeneity (Terrin, Schmid, Lau, & Olkin, 2003), but if medium-size effects are expected it is fairly appropriate (Inzlicht et al., 2015). Some other publication bias techniques were not adopted, including *p*-curve analysis and the PET technique. *p*-curve techniques consistently underperform a weight-function approach, and also involve the assumption, untenable in our case, that there are no nonsignificant studies in the meta-analytic database (McShane Bockenholt & Hansen, 2016). The PET half of PET-PEESE was also not used because it works only when the heterogeneity is low and when there is no true underlying effect, neither of which were expected in our data (Stanley, 2017). By computing and reporting several different publication bias indices, we sought to present an accurate reflection of the current state of the empirical literature with the appropriate caveats.

Of note, Vevea and Woods' (2005) weight function involves choice of direction and size for the weight function. Because both assimilation and contrast are of interest to the literature, we assumed a two-tailed publication bias. We computed both moderate and strong estimates for the weight function of publication bias, even though the inclusion of unpublished studies in the data set argues against strong publication bias.

Instances of estimated data. For two studies in the main reaction section, and four studies in the assimilation/contrast section, the standard deviations required to compute effect sizes were not reported. These standard deviations were estimated on the

[6] The sole multivariate publication bias technique is the galaxy plot implemented in the xmeta package (Chen, Hong, & Chu, 2016). Unfortunately, the associated article is not yet published and also the package can only handle two types of dependent variables at a time.

basis of standard deviations from identical measures in the same authors' own previous social comparison research, and included in the current analyses. Results without these estimated effect sizes can be accessed at the Open Science Framework's repository for this project (https://osf.io/pzrv8/) and confirm the estimations did not bias our conclusions.

Database quality control. Ten percent of the final included studies were double-coded by a student research assistant for an interrater reliability check. After discussing any disagreements, kappa values ranged from 0.83 to 0.91, except for type of the dependent variable, which had a kappa of 0.13. The entire dependent variable codes were rechecked by a student researcher and the first author until 100% agreement was reached.

Analysis program. All computations were conducted in *R*, using a variety of packages, including *metafor* (Version 1.9–9; Vichetbauer, 2010), *clubSandwich* (Version 0.2.2; Pustejovsky, 2016), and *weightr* (Version 1.1.2; Coburn & Vevea, 2017). PEESE analyses were conducted using Gervais' and Hilgard's suggested syntaxes (Gervais, 2015; Hilgard, 2015). The covariance weighting matrix was constructed as a .csv file and imported into R.

Reporting conventions. In the interests of space, cell means for nonsignificant interactions are not reported, however, full output for these analyses can be accessed via the OSF repository. Other necessary checks, such as profile plots, funnel plots, and also the full syntax and data files, can also be found on the OSF repository.

In this section, k refers to the number of effect sizes contributing to an estimate. The number of independent studies contributing to each effect size is listed in a note at the bottom of each table. Our minimum sample size for interpreting an effect size point estimate was 5 independent studies. This is consistent with standard practice (Borenstein, Hedges, Higgins, & Rothstein, 2009).

The following analyses contain a mix of multivariate and univariate analyses. To facilitate the comparison of the types of analyses, we have split the multivariate analyses and reported each dependent variable with its comparable univariate analysis.

Results

Overall Effect Size Across Studies and Dependent Variables

Overall effects without moderators for each dependent variable. The overall effect size of −0.37 ($k = 186$, $Z = -4.91$, $p < .0001$) for the multivariate model indicates contrast, as reflected by the negative effect size, rather than assimilation. This model had high heterogeneity, $Q(184) = 1683.43$, $p < .0001$.

The three uncorrected estimates suggested an effect size of −0.37 to −0.32 (see Table 3). However, both tests of publication bias (Eggers & rank-correlation) were significant, suggesting an adjusted effect size would be appropriate. The trim-fill and weight function approaches modified the univariate effect by, at most, −.05 (from −0.32 down to −0.27). Applying this back to the multivariate estimates, the adjusted overall effects were in the range of −0.37 to −0.32. As will be seen throughout, the PEESE estimates often varied widely from the other estimates (see the preceding text for earlier caveats about PEESE).

Table 3
Estimates of Overall Effect Size

Estimate type	*M*	95% CI
Uncorrected estimates		
Multivariate (covariance known)	−.37	[−.52, −.22]
Robust (sandwich)	−.37	[−.52, −.22]
Univariate random effects	−.32	[−.44, −.20]
Publication bias tests		
Egger's test	$Z = 3.24$, $p = .001$	
Rank-correlation test	$\tau = .13$, $p = .007$	
Publication-bias corrected estimates		
Trim & fill	−.32	[−.45, −.20]
Weight function (moderate 2 tail)	−.29	
Weight function (strong 2 tail)	−.27	
PEESE	−.60	[−.67, −.54]

Note. The Vevea and Woods (2005) weight function does not provide CIs.

Overall effect for each dependent variable. There was a significant multivariate effect for dependent variable type, $Q_m(4) = 28.80$, $p = .000$. As can be seen in Table 4, ability and performance satisfaction measures showed contrast, while behavior, affect and self-esteem showed no overall effects. The variety of estimation methods and sensitivity analysis suggested a range of likely values for each dependent variable but in all cases ability and performance satisfaction remained contrastive while behavior, affect and self-esteem remained nonsignificant.

Of the two contrastive dependent variables, ability had a significant Egger's test and the range of estimates was between −0.27 (Vevea & Woods' severe publication bias) and −0.43 (multivariate), all of which are significant. Performance satisfaction measures also showed strong contrast, with an estimate between −1.4 to −1.0. Results for self-esteem overlapped with 0.0 and were also associated with publication bias— a marginally significant Egger's test and a trim-and-fill adjustment reduced to the effect 0.01 from the original value of 0.16.

Moderator Analysis

(Dis)similarity inductions. There was a main effect of (dis)similarity inductions ($Z = -4.66$, $p < .0001$), but no effect of dependent variable type, $Q_m(4) = 5.93$, $p = .205$, nor an interaction, $Q_m(2) = 1.48$, $p = .476$. There was also no significant variance at the study level once the (dis)similarity moderator was included in the model (see online profile plots). Hence, we examined the main effect of (dis)similarity by dropping the dependent variable type and dropping the random effect for study code. The moderator was again significant ($Z = -4.69$, $p < .0001$), suggesting similarity and dissimilarity have different effects.

As seen in Table 5, the dissimilarity estimates ranged from −.27 to −.44. The results of the trim-and-fill method suggest there is a modest to moderate effect, and the weight function estimates suggested the same, but the CI borders 0.0. There were no indications, however, of publication bias in the Eggers and/or rank correlation tests. Contrast of small-to-moderate magnitude seems to follow dissimilarity priming.

The similarity studies showed indications of publication bias, with both bias tests reaching significance. The adjusted effects values range from 0.02 (PEESE method) to 0.39 (moderate version of Vevea &

Table 4
Effect Sizes for Each Type of Dependent Variables

Analysis	Ability (k = 102)		Affect (k = 34)		Self-esteem (k = 21)		Behavior (k = 19)		Performance satisfaction (k = 9)	
	d	95% CI	d	95% CI	d	95% CI	d	95% CI	d	95% CI
Uncorrected estimates										
Multivariate	−.34	[−.50, −.17]	−.17	[−.45, .11]	−.16	[−.52, .19]	−.13	[−.51, .24]	−1.47	[−1.99, −.95]
Univariate	−.33	[−.49, −.17]	−.17	[−.45, .11]	−.17	[−.52, .18]	−.13	[−.50, .24]	−1.52	[−2.04, −1.00]
Robust	−.34	[−.49, −.18]	−.17	[−.49, .15]	−.16	[−.52, .19]	−.13	[−.46, .19]	−1.47	[−2.21, −.74]
Publication bias tests										
Egger's test	Z = 2.40, p = .02		Z = .94, p = .35		Z = 1.68, p = .09		Z = 1.38, p = .17		Z = .01, p = .99	
Rank–correlation test	τ = .10, p = .12		τ = .16, p = .19		τ = .24, p = .14		τ = .26, p = .12		τ = .11, p = .76	
Publication–bias corrected estimates										
Trim & fill	−.33	[−.49, −.17]	−.17	[−.45, .11]	.01	[−.33, .36]	−.14	[−.42, .14]	−1.62	[−2.29, −.94]
Vevea & Woods (moderate)	−.29		−.15		−.15		x		−1.47	
Vevea & Woods (severe)	−.27		−.14		−.14		−.11		−1.44	
PEESE	−.71	[−.96, −.46]	−.14	[−.54, .26]	−.67	[−1.35, .01]	−.68	[−1.34, −.02]	−1.07	[−3.52, 1.38]

Note. Study sizes for each column are 57, 25, 13, 16, and 9.

Woods). Estimates based on PEESE suggest that similarity priming has no effect (see Table 5). However, across all of our analyses PEESE's estimates are the least consistent with the other estimates and vary widely. Moreover, other simulations (Inzlicht, Gervais, & Berkman, 2015; Hilgard, 2015) suggest caution is advised in the interpretation of PEESE. Balancing these considerations, the available data suggest that similarity priming may lead to weak assimilation after social comparison. In light of the attention paid to the SAM and to similarity priming in reviews and textbooks (e.g., Mussweiler, 2003; Suls, Martin, & Wheeler, 2002), however, the case for similarity priming to induce assimilation requires additional study. We urge that these studies include control groups and larger samples than has been the case in the past.

Local/distant target. The overall local/distant target by dependent variable model was significant, $Q_m(9) = 36.03$, $p < .0001$. There was a marginal effect for the moderator ($Z = -1.94$, $p = .053$), significant effects for dependent variable type, $Q_m(4) = 10.42$, $p = .034$ and the interaction, $Q_m(4) = 10.30$, $p = .036$. As the interaction was significant, we examined each dependent variable separately.

Ability. The strong contrast effect of local comparisons is shown in all estimates (0.7–0.8), with little evidence of publication bias (see Table 6). Even (the conservative) PEESE places the value at −0.6. The influence of a distant target is more variable (from −0.23 to −0.4), but is non-significant as the CI includes zero in all uncorrected estimates (see Table 6, rows 1–3) and the Egger's test indicates there is some bias.

Affect. Affect tended to favor contrast after local comparisons (see Table 7), with most estimates ranging from −0.75 to −0.67, but the CI's included zero for three estimators. Distant comparisons had no systematic influence on affect.

Behavior. There were no discernible behavioral effects following local or distant comparisons (see Table 8). The magnitudes of the mean effect sizes were consistent but the confidence intervals were wide.

Table 5
Effects of (Dis)Similarity Manipulation/Priming

Effect size type	Similarity (k = 16)		Dissimilarity (k = 37)	
	Mean ES (d)	95% CI	Mean ES (d)	95% CI
Uncorrected estimates				
Multivariate	.52	[.18, .85]	−.44	[−.65, −.22]
Univariate	.51	[.16, .86]	−.44	[−.67, −.21]
Robust	.52	[.20, .83]	−.44	[−.82, −.05]
Publication bias tests				
Egger's	Z = 2.11, p = .04		Z = −.78, p = .43	
Rank correlation	τ = .39, p = .04		τ = −.10, p = .39	
Corrected estimates				
Trim and fill	.51	[.19, .82]	−.27	[−.52, −.02]
Weight function (moderate)	.44		−.38	
Weight function (severe)	.39		−.34	
PEESE	−.02	[−.41, .45]	−.36	[−.72, .00]

Note. ES = Effect size. Study sizes for columns are 10 and 26.

Table 6
Effects of Target Immediacy on Ability Estimates

Ability	Local (k = 8)		Distant (k = 94)	
Type	Mean ES (d)	95% CI	Mean ES (d)	95% CI
Uncorrected estimates				
Multivariate	−.84	[−1.38, −.30]	−.29	[−.45, −.12]
Univariate	−.84	[−1.38, −.29]	−.28	[−.45, −.12]
Robust	−.84	[−1.29, −.4]	−.29	[−.76, .19]
Publication bias tests				
Egger's test	Z = .10, p = .92		Z = 2.11, p = .04	
Rank–correlation test	τ = −.36, p = .28		τ = .10, p = .16	
Publication–bias corrected estimates				
Trim & fill	−.95	[−1.41, −.50]	−.28	[−.45, −.11]
Henmi & Copas	−.73	[−1.49, .03]	−.40	[−.61, −.20]
Vevea & Woods (moderate)	−.79		−.25	
Vevea and Woods (severe)	−.76		−.23	
PEESE	−.64	[−1.32, .04]	−.70	[−.99, −.41]

Note. ES = Effect size. Study sizes for columns are 8 and 49.

Self-esteem and performance satisfaction. There were only four self-esteem effect sizes in the local condition and the distant condition was nonsignificant. The performance satisfaction variable also had low ks (of 4 and 5).

Dimension novelty. The overall model was significant, $Q_m(8) = 41.41$, $p < .0001$, with a significant main effect for DV, $Q_m(4) = 28.58$, $p < .0001$, a marginal effect for novelty ($Z = -1.85$, $p = .064$), and a significant interaction, $Q_m(3) = 16.14$, $p = .001$. Tables for each of the dependent variables are provided.

Ability. Comparisons on novel dimensions yielded strong contrast, from −0.5 (weight function) up to −0.6 (most methods). The two publication bias tests were nonsignificant and the trim-and-fill estimate returned a larger (not smaller) estimate than the uncorrected techniques (see Table 9). Comparisons on known dimensions had a weak effect on ability estimates. All estimation approaches converged on −0.25 although Egger's test was significant. The publication bias corrections indicated the range of this effect ranges between −0.2 (weight function) to −0.5 (trim-and-fill).

Affect. Comparisons on known dimensions had no effect on affect (see Table 10), but comparisons on novel dimensions led to strong contrast of affect. The estimates of effect size here were relatively consistent, ranging from −0.77 to −0.74.

Behavior. Neither known nor novel dimensions produced any discernible effect on behavior (see Table 11). In all cases, the confidence intervals encompassed zero.

Self-esteem and performance satisfaction. Self-esteem had no studies with novel dimensions, and performance satisfaction had low ks.

Varying self versus standard. The overall model was significant, $Q_m(9) = 39.30$, $p < .0001$, with main effects for DV, $Q_m(4) = 19.08$, $p = .001$, the moderator ($Z = 2.25$, $p = .025$), and a marginally significant interaction, $Q_m(4) = 9.35$, $p = .053$. We analyzed the interaction.

Ability. Varying the participant's standing led to strong contrast (see Table 12), with values ranging from −0.75 to −0.65. The publication bias tests were not significant but the weight function returned a slightly attenuated effect size. When only the standard was varied, there was no overall effect. Although the

Table 7
Effects of Target Immediacy on Affect

Type	Local (k = 8)		Distant (k = 26)	
	M ES (d)	95% CI	M ES (d)	95% CI
Uncorrected estimates				
Multivariate	−.73	[−1.27, −.19]	.03	[−.29, .34]
Univariate	−.75	[−1.30, −.19]	.03	[−.29, .35]
Robust	−.73	[−1.23, −.23]	.03	[−.59, .64]
Publication bias tests				
Egger's test	Z = −1.27, p = .20		Z = .60, p = .55	
Rank–correlation test	τ = −.29, p = .40		τ = .13, p = .36	
Publication–bias corrected estimates				
Trim and fill	−.67	[−.91, −.43]	.03	[−.30, .36]
Vevea and Woods (moderate)	−.70		.02	
Vevea & Woods (severe)	−.67		.02	
PEESE	−.32	[−1.19, .55]	.02	[−.43, .47]

Note. ES = Effect size. Study sizes for columns are 7 and 19.

Table 8
Effects of Target Immediacy on Behavior

Type	Local (k = 7)		Distant (k = 12)	
	M ES (d)	95% CI	M ES (d)	95% CI
Uncorrected estimates				
Multivariate	.15	[−.45, .74]	−.30	[−.75, .16]
Univariate	.15	[−.46, .75]	−.30	[−.76, .17]
Robust	.15	[−.47, .76]	−.30	[−1.02, .42]
Publication bias tests				
Egger's test	Z = .07, p = .94		Z = 1.83, p = .07	
Rank–correlation test	τ = .33, p = .38		τ = .33, p = .15	
Publication–bias corrected estimates				
Trim and fill	.15	[−.27, .56]	−.32	[−.66, .03]
Vevea and Woods (moderate)	.12		−.28	
Vevea and Woods (severe)	.10		−.25	
PEESE	.09	[−1.29, 1.48]	−1.03	[−1.77, −.30]

Note. ES = Effect size. Study sizes for columns are 7 and 9.

standard-only effect size was nonsignificant, this was an instance in which the publication bias tests were significant. However, only the weight function method corrected the results in a way that attenuated the effect size due to publication bias (both trim-and-fill and PEESE increased the estimate of the effect size).

Affect. Varying the participants' standing led to strong contrast effects (see Table 13). Although the tests for publication bias were nonsignificant, the adjustment methods suggest that the uncorrected estimate of −0.83 could be adjusted down to −0.75 (trim-and-fill sets it at −0.71, weight function at −0.77), and, according to PEESE, −0.65. There were no effects when only the standard was varied.

Self-esteem, behavior and performance satisfaction. There was an insufficient number of studies representing the manipulation of self/standard feedback to test for effects on self-esteem, behavior or performance satisfaction.

Explicit/implicit inductions. The overall model of explicit/implicit task by dependent variable was significant, $Q_m(9) = 27.55$, $p = .001$, indicating there was an overall average effect after adjusting for other dependent variables. There were no significant effects for any moderators. There was a marginal effect for the explicit/implicit induction ($Z = -1.86$, $p = .063$), but no effect for dependent variable type, $Q_m(4) = 2.08$, $p = .721$ nor the interaction, $Q_m(4) = 2.27$, $p = .686$. Dropping the interaction and dependent variable type from the model, the effect of the moderator remained only marginally significant ($Z = -1.73$, $p = .082$). A table with the specific effect size estimates is available on the OSF repository.

Exemplars. Although the overall multivariate model was significant, $Q_m(6) = 28.02$, $p < .0001$ there was no significant moderator effect ($Z = -1.44$, $p = .149$) and the only interaction involved performance satisfaction with very low k. Univariate results were the same so this moderator was not analyzed further.

Objective/subjective dependent variables. The objective/subjective distinction was only represented in those studies reporting ability estimates. The overall regression of dependent variable type and objective/subjective type was not significant, $Q_m(1) = .16$, $p = .690$.

Self-other/other-self direction. The overall regression was not significant, $Q_m(6) = 3.23$, $p = .778$. Consequently, there were no further analyses conducted.

Table 9
Effect of Dimension Novelty on Ability Estimates

Type	Known (k = 80)		Novel (k = 22)	
	M ES (d)	95% CI	M ES (d)	95% CI
Uncorrected estimates				
Multivariate	−.25	[−.43, −.08]	−.60	[−.92, −.28]
Univariate	−.25	[−.42, −.07]	−.60	[−.92, −.28]
Robust	−.25	[−.44, −.06]	−.60	[−.94, −.25]
Publication bias tests				
Egger's test	Z = 1.99, p = .047		Z = −.12, p = .91	
Rank–correlation test	τ = .07, p = .38		τ = −.07, p = .66	
Publication–bias corrected estimates				
Trim & fill	−.51	[−.71, −.30]	−.78	[−1.08, −.49]
Vevea and Woods (moderate)	−.22		−.55	
Vevea and Woods (severe)	−.20		−.51	
PEESE	−.74	[−1.08, −.40]	−.56	[−.96, −.16]

Note. ES = Effect size. Study sizes for columns are 39 and 18.

Table 10
Effects of Dimension Novelty on Affect

Type	Known ($k = 26$)		Novel ($k = 7$)	
	M ES (*d*)	95% CI	*M* ES (*d*)	95% CI
Uncorrected estimates				
Multivariate	−.02	[−.34, .29]	−.77	[−1.33, −.21]
Univariate	−.01	[−.33, .30]	−.77	[−1.34, −.20]
Robust	−.02	[−.40, .36]	−.77	[−1.28, −.26]
Publication bias tests				
Egger's test	$Z = .41, p = .70$		$Z = -.70, p = .48$	
Rank–correlation test	$\tau = .05, p = .73$		$\tau = -.05, p = 1$	
Publication–bias corrected estimates				
Trim and fill	−.01	[−.35, .32]	−.75	[−.91, −.58]
Vevea and Woods (moderate)	−.01		−.74	
Vevea and Woods (severe)	−.01		−.74	
PEESE	.19	[−.28, .67]	−.66	[−1.11, −.21]

Note. ES = Effect size. A *p*–value of 1 is possible for rank–correlation tests with low *k*. Study sizes for columns are 18 and 7.

Summary of Reaction Results

Overall, contrast was, by far, the dominant response to social comparison. Ability estimates were the most strongly affected. This is not so surprising as social comparison theory was originally developed to explain ability self-evaluations. Moderator analyses show that contrast is stronger when the comparison involves varying participants' standing for ability (−0.75 to −0.65) and affect (−0.83 to −0.65). Also, contrast is stronger for ability estimates (−0.5 to −0.6) and affect (−0.6 to −0.7) when the personal attribute is novel.

Dissimilarity priming was associated with contrast (−0.44 to −0.27; no publication bias). Similarity priming, at best, showed very weak assimilation effects, depending on the publication bias estimator. (In fact, PEESE, which is most inconsistent with other estimators in simulation work [Inzlicht et al., 2015], would declare there is no similarity priming effect.)

Across reaction studies (without control groups) the weight function selection approach returned consistently more conservative estimates and seemed to be in agreement with the trim and fill test, about which some metaanalysts have had reservations. Inzlicht et al. (2015), however, have observed that trim and fill can work well for medum effect sizes (i.e., our situation). Finally, multivariate and univariate techniques, for the most part, did not yield different results. This may be because only 20% of studies had more than one DV.

Meta-Analysis of Reaction Studies With Control Groups

The preceding section assessed the prevalence of contrast (compared with assimilation) but was unable to separate contrast and assimilation into upward and downward components. To ascertain whether upward (or downward) assimilation and upward (or downward) contrast occur, it is necessary to test whether effects of an upward (or downward) comparison target differ from a control condition, as illustrated in Figure 1.

Although use of control conditions can be found in the literature, they are less common than the upward versus downward design (reviewed earlier). Across all dependent variables, the control conditions used were either no comparison ($k = 95$), a lateral

Table 11
Effects of Dimension Novelty on Behavior

Type	Known ($k = 12$)		Novel ($k = 7$)	
	M ES (*d*)	95% CI	*M* ES (*d*)	95% CI
Uncorrected estimates				
Multivariate	−.29	[−.75, .16]	.11	[−.47, .69]
Univariate	−.28	[−.74, .18]	.11	[−.48, .70]
Robust	−.29	[−.68, .10]	.11	[−.58, .8]
Publication bias tests				
Egger's test	$Z = .96, p = .34$		$Z = 2.08, p = .04$	
Rank–correlation test	$\tau = .27, p = .25$		$\tau = .43, p = .24$	
Publication–bias corrected estimates				
Trim and fill	−.29	[−.63, .05]	.10	[−.38, .57]
Vevea and Woods (moderate)	−.25		.07	
Vevea and Woods (severe)	−.23		.06	
PEESE	−.82	[−1.76, .12]	−.85	[−2.00, .30]

Note. ES = Effect size. Study sizes for columns are 9 and 7.

Table 12
Effect of Varying Self Versus Standard on Ability Estimates

Type	Self (k = 15)		Standard (k = 87)	
	M ES (*d*)	95% CI	*M* ES (*d*)	95% CI
Uncorrected estimates				
Multivariate	−.75	[−1.14, −.36]	−.26	[−.43, −.09]
Univariate	−.73	[−1.13, −.34]	−.26	[−.43, −.09]
Robust	−.75	[−1.05, −.45]	−.26	[−.60, .09]
Publication bias tests				
Egger's test	$Z = -.30, p = .77$		$Z = 2.12, p = .03$	
Rank–correlation test	$\tau = -.23, p = .23$		$\tau = .10, p = .16$	
Publication–bias corrected estimates				
Trim and fill	−.83	[−1.12, −.54]	−.51	[−.70, −.31]
Vevea and Woods (moderate)	−.68		−.23	
Vevea and Woods (severe)	−.64		−.21	
PEESE	−.65	[−1.28, −.03]	−.68	[−.96, −.40]

Note. ES = Effect size. Study sizes for columns are 15 and 44.

comparison ($k = 33$), or a nonsocial comparison (e.g., to a product, $k = 2$).

Analysis of studies with control groups permitted the examination of two questions. The first was the relative effect magnitude of upward and downward comparisons. In the extreme case, upward (or downward) comparisons may be indistinguishable in their effects from control conditions. This pattern would be masked in the analyses reported in the previous section because there were no control conditions in the reported studies. To investigate the possibility that upward comparisons (vs. control conditions) were stronger than downward comparisons (vs. control conditions), the main effect for comparison direction was tested in a multivariate regression model.

The second question was whether any instances of assimilation in the preceding analysis had been overlooked (recalling that contrast was the predominate outcome). Some individual conditions might have produced assimilation, even without yielding a significant main effect or interaction. To detect assimilation, we used backward elimination to compute a parsimonious moderator model, and then looked for assimilation in the individual cell means.

Method

All effect sizes that used a control condition were extracted from the main database. These were stacked into one file, and an index was created to note whether the effect size represented an upward comparison or a downward comparison. Our final control groups database had 130 effect sizes from 39 articles (see Appendix D in the online supplemental material). Some studies had multiple dependent variables, hence a multivariate covariance matrix was computed to allow proper estimation of the composite effect sizes. As was the case with the reaction studies without control groups, a reasonable number of dependent variable correlations (36 out of 59) were recovered from study authors. The remaining correlations were estimated using existing data.

ESs were coded such that positive effect sizes reflected assimilation and negative effect sizes reflected contrast. This was true in both upward and downward cases. To clarify, a positive effect size in an upward versus control contrast indicates upward assimilation, and a negative effect size indicates upward contrast. A positive effect size in a downward versus control contrast indicates downward assimilation, and a negative effect size indicates downward contrast, as shown in Figure 1. All other coding used the scheme reported in the prior section on reaction studies without control groups.

Analyses were again conducted using the same R packages as the prior section, including *metafor, weightr* and *clubSandwich*. Coding for the PEESE estimates was again borrowed from scripts by Will Gervais and Joe Hilgard.

Table 13
Effects of Varying Self Versus Standard on Affect

Type	Self (k = 8)		Standard (k = 26)	
	M ES (*d*)	95% CI	*M* ES (*d*)	95% CI
Uncorrected estimates				
Multivariate	−.83	[−1.37, −.30]	.06	[−.26, .38]
Univariate	−.82	[−1.36, −.28]	.06	[−.26, .38]
Robust	−.83	[−1.16, −.51]	.06	[−.43, .54]
Publication bias tests				
Egger's test	$Z = -1.08, p = .28$		$Z = .34, p = .73$	
Rank–correlation test	$\tau = -.14, p = .72$		$\tau = .06, p = .66$	
Publication–bias corrected estimates				
Trim and fill	−.71	[−.88, −.55]	.06	[−.26, .38]
Vevea and Woods (moderate)	−.77		.05	
Vevea and Woods (severe)	−.77		.04	
PEESE	−.64	[−1.05, −.23]	.18	[−.27, .63]

Note. ES = Effect size. Study sizes for columns are 8 and 18.

Results

First, a model to test whether dependent variable type had an effect was conducted. There was no overall effect of DV type, $Q_m(4) = 3.25$, $p = .517$, so dependent variable type was dropped from further consideration.

The multivariate test of direction was computed and it was not significant, $Q_m(1) = 0.48$, $p = .491$. There was no evidence that one comparison direction was stronger than the other in its effects.

As seen in Table 14, the downward-versus-control studies are associated with an estimate of the true effect between −0.16 and −0.27, but this should be considered in light of possible publication bias (the corrected estimates did not concur). The upward-versus-control condition estimates ranged from −0.14 to −0.18. The CI's for the univariate and robust tests contained 0.0, but the corrected estimates did not. In light of the nonsignificant multivariate results, one (cautious) conclusion is that there is no overall difference in the magnitude of the effects for reactions to upward and downward comparisons. More critically, however, there is a modest indication that downward comparisons produce positive effects and upward comparisons produce (with less certainty) negative effects versus control groups. Provisonally, the meta-analysis of studies with control groups indicates effects are small, they are only contrastive, and by implication exposure to the comparison target is "pushing" self-evaluations away; responses were not "pulled" toward the target.

In order to look for any instances of assimilation that were overlooked, a model including all the moderators and their interactions with direction was computed. A single factor, the interaction between direction and self-standard variation, was significant. We ran a reduced model with just these main effects and the interaction.

The overall model was significant, $Q_m(3) = 18.26$, $p = .000$. The mean effect for each level of moderator is shown in Table 15. There were no cells that showed an overall assimilative effect, with most cells showing contrast. These results are consistent with the meta-analytic results presented earlier for the reaction studies that lacked control groups. For the purposes of completeness, if the standard was varied and the comparison was with a local target made in a downward direction, no contrast was exhibited. Given that none of these effects were of any particular theoretical interest, we omitted the publication bias indicators, although they are available online.

We also ran a model with (dis)similarity (which had been removed from the large model because only a subset of studies have the (dis)similarity moderator). This model was not significant, $Q_m(3) = 1.54$, $p = .673$. This difference from the reactions for studies without control groups is not surprising given the smaller number of studies in the present analysis and that we had split the sample with respect to one of the weaker moderators (upward-control and downward-control).

Summary of Reaction Results From Studies With Control Groups

There was no evidence of an overall effect for upward comparisons to produce a larger contrast effect than did downward comparisons. Also there was no evidence of assimilation. There are indications that upward and downward comparisons "push" responses away from target standing.

Discussion

These are the first meta-analyses devoted to two long-standing issues in social comparison: (1) selection of a comparison target and (2) reactions to comparison. The two issues will be discussed separately before commenting upon their relationship to one another. In discussing selection of a comparison target, particular attention will be paid to the context of the comparison selection, whether the comparer is threatened or not, whether the study takes place in the laboratory or the field, and whether there is a lateral choice in addition to upward and downward choices. For reactions to comparison, the focus will be on variables that might moderate the degree of assimilation toward or contrast away from the comparison target. In addition to these substantive issues, the meta-analytic findings will be discussed with appropriate caveats in view of recent concerns about publication bias and threats to interpretation.

Table 14
Effects of Direction on Effect Size

Effect size type	Up–control ($k = 79$)		Down–control ($k = 51$)	
	M ES (*d*)	95% CI	*M* ES (*d*)	95% CI
Uncorrected estimates				
Multivariate	−.18	[−.29, −.07]	−.22	[−.33, −.11]
Univariate	−.15	[−.32, .03]	−.27	[−.45, −.08]
Robust	−.18	[−.39, .03]	−.24	[−.43, −.01]
Publication bias tests				
Egger's	$Z = 1.07, p = .28$		$Z = -1.72, p = .08$	
Rank–correlation	$\tau = .05, p = .51$		$\tau = -.06, p = .57$	
Corrected estimates				
Trim and fill	−.18	[−.30, −.06]	−.20	[−.40, −.01]
Weight function (moderate)	−.16		−.18	
Weight function (severe)	−.14		−.16	
PEESE	−.32	[−.51, −.13]	.26	[−.13, .65]

Note. ES = Effect size. Study sizes for columns are 40 and 27.

Table 15
Cell Means for Direction X Immediacy X Self–Standard Interaction

DV type	k	Self-varied D	95% CI	k	Standard varied d	95% CI
Up-local	8	−.61	[−1.43, −.03]	9	−.46	[−1.17, −.01]
Down-local	8	−.65	[−1.70, −.23]	2	.03	[−.95, .36]
Up-distant	7	−.52	[−1.17, .03]	54	−.36	[−.76, −.10]
Down-distant	7	−1.07	[−1.79, −.56]	33	−.40	[−.85, −.15]

Note. DV = dependent variable. Study sizes by column are (4, 4, 5, 5) and (6, 2, 24, 16).

Selection Studies

The analysis of selection studies showed a somewhat mixed pattern when there was no possibility of a lateral comparison. When participants were limited to either an upward or downward choice, laboratory and field studies produced different results. In laboratory studies, threat reduced the frequency of upward comparisons. However, even under threat, choices were still upward (74%). In field recall studies, on the other hand, threat increased the frequency of upward choices, but this result is based on a very small number of studies. There is no evidence in either laboratory or field recall studies for threat leading to downward comparison, contrary to downward comparison theory's selection prediction.

When participants were allowed the possibility of a lateral comparison, choices among the three options were less differentiated than when only two options were provided. Having a lateral choice available reduced the differences among choice preferences, but upward selections still were preferred. Also, when lateral comparison was an option, neither threat nor setting moderated selections (in contrast to the two-choice selection studies). The smaller number of studies offering a third option may be contributing to the loss of the interactions. A research-worthy hypothesis is that the lateral choice presents the opportunity to share a "similar-fate" (Darley & Aronson, 1966; Schachter, 1959; Wills, 1981), which competes with the striving to identify with a person of higher standing. Presenting a lateral option may make the two (conflicting) motives salient, thereby complicating the selection process for the participant.

Reaction Studies

Most social comparison research has not included a control group or pretest and has simply compared the reaction to an upward comparison with the reaction to a downward comparison. We analyzed such studies separately from those that did have a control group. ES was coded so that a positive ES indicated upward or downward assimilation, and a negative ES indicated upward or downward contrast. Across all outcomes, contrast was the predominate outcome ($d = -0.37$); publication bias estimates only lowered the magnitude slightly, suggesting that the default response to comparison is contrast. Unfortunately, because of the lack of control groups, determining the magnitude of upward contrast (self-evaluation decreasing as a result of upward comparison) relative to the magnitude of downward contrast (self-evaluation increasing as a result of downward comparison) is not possible for these studies.

There were overall effects of comparison on ability assessments and performance satisfaction, consistently showing contrast, −0.43 to −0.27, and −1.4 to 1.0, respectively, but no overall effects on the other dependent variables.

Target immediacy (Buckingham & Alicke, 2002; Zell & Alicke, 2010) had a significant effect. Although local targets and distant targets both resulted in contrast for ability, the effect was larger (with no evidence of publication bias or mis-estimation) for the local targets (−0.6 to −0.7) than for the distant targets (−0.2 to −0.4), which exhibited some bias. Affect also tended to be more elevated after knowing that you out-performed a person sitting next to you than knowing that you outperformed a person from the previous year's class (notwithstanding some indication of publication bias). Zell and Alicke explain local dominance as being due to humans having evolved in small groups and having been socialized in small groups, both leading to a focus on local comparisons. Social class is now being studied as a rank-related construct, and Norton (2013) has argued for the greater importance of local ranks over population ranks. People, he says, are painfully aware of their local rank and surprisingly unaware of their population rank.

Known-novel dimensions emerged in an overall analysis and specifically with respect to ability and affect. When the personal attribute is a relatively known dimension, such as intelligence or strength, there is a weak contrast effect ($d = -0.25$) for ability judgments, but that effect is stronger (−0.5 to −0.6 and with no indication of publication bias) for a relatively unknown dimension, such as cognitive flexibility. This is not surprising, as the known dimensions are more firmly anchored in past experiences. The effect on affect was similar, with unknown or novel dimensions showing greater negative moods subsequent to comparison; for familiar attributes, comparison had no influence on affect.

One can also distinguish between experimental situations in which feedback about the self is manipulated versus situations in which feedback about the standard is manipulated. When feedback about the self was manipulated, there was a sizable impact (approximately −0.6 to −0.8) on ability and affect. Manipulating the standard, however, had no discernible influence on ability or affect. People have a view of their ability levels, and getting feedback contradicting that view has an effect.

When people were made to feel dissimilar or primed with dissimilarity, they contrasted their self-evaluations away from the target. The magnitude of contrast, however, is modest, −0.44 to −0.27 (depending on the estimator), but with no indication of publication bias.

There was one notable exception to the bulk of results showing contrast. When participants were made to feel similar to the comparison target or were primed with thoughts of similarity, there some indication of weak assimilation (+0.02 to +0.39) of their self-evaluations toward the target. However, we are not entirely confident, as these results were based on a modest number of studies, and the two publication bias tests and PEESE suggest a cautious interpretation.

These results provide partial support for SAM (Mussweiler (2003; Mussweiler & Strack, 2000). Dissimilarity priming does apparently shift evaluations from the comparison target. Both SAM and Collins (1996, 2000) also predicted that upward assimilation would occur when the comparer felt similar to the target or similarity was primed. Support for upward assimilation following similarity priming was weak, however. Nonetheless, it is notable that priming similarity revealed the only evidence for assimilation here; otherwise, contrast seems to be the default response.

One can use as a comparison standard an exemplar, such as Einstein, or an unspecified member of the same category, such as college professor. Dijksterhuis et al. (1998) argued that only the exemplars produced comparison and thus contrast, and that nonexemplars acted only as primes and did not evoke comparison. There was no moderation of the effect size by whether exemplars or nonexemplars were used, and hence we found no support for Dijksterhuis et al.'s argument in the current meta-analysis.

There was no difference between objective and subjective dependent variables. Although Mussweiler (2003) has argued that his SAM should work best with objective dependent variables (as noted in the introduction), we found no evidence for a difference in results between the two types of variables. This suggests to us that there may be "cross-talk" between the language people use to assess their abilities (e.g., "very strong," a "7") and higher level mental representations that, according to Mussweiler's theory, are reflected by objective variables (e.g., "I can do 100 push-ups"; Suls & Wheeler, 2017).

There were also no effects for implicit/explicit inductions or for self-versus-other focus. Perhaps the limited number of studies of these factors restricted the power to detect effects.

The analysis of the control group studies represented a smaller sample, which no doubt restricted the ability to detect different effects across outcomes and moderators. The notable finding was upward and downward targets have approximately the same magnitude of impact, ranging from −0.14 to −0.27, with downward comparison associated with positive and upward comparison with negative effects on evaluations—again indicating that contrast appears to be the most common consequence of social comparison.

Reconciling Social Comparison Choice and Reaction

The great conundrum of social comparison is why people choose to compare upward when the most likely result is a self-deflating contrast. It would seem more logical to make downward comparisons in order to feel better about oneself, and it is that logic that perpetuates a popular acceptance of downward comparison theory's second hypothesis. But we think that people do not adequately anticipate the self-deflating contrast, or that the contrasts will be outweighed by other benefits. They think that they will be "almost as good as the very good ones" (Wheeler, 1966), that they will be able to assimilate themselves to a higher level, that they will learn the secrets of being better, and so forth. Women buy fashion magazines that force them into self-deflating contrasts and men do exactly the same thing with fashion and body building magazines (e.g., Grabe, Ward, & Hyde, 2008; Myers & Crowther, 2009). And indeed the magazines may well be worth the candle because of the other benefits that come along with the self-deflating comparisons.

There is also the persistent Western belief that we are better than the average peer. We are better drivers, better lovers, and better parents than the average person. The better-than-average effect (BTAE) seems to be due to some factors that are independent of social comparison. Alicke and Govorun (2005) explain it as an "automatic tendency to assimilate positively-evaluated social objects toward ideal trait conceptions" (p. 99; see also Brown, 2012). If you were asked to indicate how kind you are relative to the average person, and if you thought you were generally kind, you would assimilate your view of your kindness to the ideal point on the kindness scale, which would make you kinder than the average person. More cognitive explanations may also contribute to the BTAE, such as the selective recruitment of information about the self because it is more accessible (Dunning, Meyerowitz, & Holzberg, 1989) or applying greater weight to own characteristics in judgment (Chambers & Windschitl, 2004). Regardless of the BTAE's specific origins, if one goes through life believing and expecting to be better than the average person, upward comparison would seem to be a natural consequence, despite the occasional rude evidence that someone else is better than you.

Conclusions

Although the limited number of studies including no-comparison control groups, sparse representation of some dependent variables, and instances of publication bias leave some questions unresolved, these meta-analyses offer several conclusions. Given the choice, the predominant tendency is to compare upward; threat may temper it, but downward comparison is not a dominant choice. Comparisons are more potent when they are local, involve unknown dimensions, or manipulate the self (rather than a standard).

The common response to comparison is contrast: people increase their self-evaluations after downward comparison and decrease their self-evaluations after upward comparisons. These effects appear to be approximately equivalent in magnitude. Assimilation, which has been stressed in recent research, appears not to be the default; it requires special priming. Even then, this review indicates those effects are weak and in need of further study.

To understand the seeming conflict between the results from the selection and response literatures, we speculate that people presume they are good, but this coexists with "a congenital uncertainty" (de Botton, 2004, p. 8), so they look upward to confirm their closeness to the "better ones," which often leads, alas, to self-deflation. Although Festinger did not anticipate all of our meta-analytic findings, he recognized that strivings to be similar to others conflicted with the unidirectional drive upward so that social quiescence with respect to abilities would never be attained. We witness this every day in domains as different as income, physical attractiveness, personality and aptitude. Social comparison theory has been extraordinarily generative, not because of the clarity or brilliance of the theory, but because it deals with an inescapable aspect of human life. The theory, as shown by the many unanswered questions in these meta-analyses, needs a new generation of scientists who appreciate the importance of control groups and who recognize the inevitability of social comparison.

References

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology, 49,* 1621–1630. http://dx.doi.org/10.1037/0022-3514.49.6.1621

Alicke, M. D., & Govorun, P. (2005). The better-than-average effect. In M. D. Alicke, D. A. Dunning, & J. I. Krueger (Eds.), *The self in social judgment* (pp. 85–108). New York, NY: Taylor & Francis.

Allan, S., & Gilbert, P. (1995). A social comparison scale: Psychometric properties and relationship to psychopathology. *Personality and Individual Differences, 19,* 293–299. http://dx.doi.org/10.1016/0191-8869(95)00086-L

Arigo, D., Suls, J. M., & Smyth, J. M. (2014). Social comparisons and chronic illness: Research synthesis and clinical implications. *Health Psychology Review, 8,* 154–214. http://dx.doi.org/10.1080/17437199.2011.634572

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50,* 1088–1101. http://dx.doi.org/10.2307/2533446

Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology, 60,* 485–499. http://dx.doi.org/10.1037/0022-3514.60.4.485

Blanton, H., Crocker, J., & Miller, D. T. (2000). The effects of in-group versus out-group social comparison on self-esteem in the context of a negative stereotype. *Journal of Experimental Social Psychology, 36,* 519–530.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. New York, NY: Wiley. http://dx.doi.org/10.1002/9780470743386

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36,* 129–148. http://dx.doi.org/10.1037/0003-066X.36.2.129

Bower, G. H. (1991). Mood congruity of social judgments. In J. P. Forgas (Ed.), *Emotion and social judgments* (pp. 31–53). Oxford, UK: Pergamon Press.

Brewer, M. B., & Weber, J. G. (1994). Self-evaluation effects of interpersonal versus intergroup social comparison. *Journal of Personality and Social Psychology, 66,* 268–275. http://dx.doi.org/10.1037/0022-3514.66.2.268

Brown, J. D. (2012). Understanding the better than average effect: Motives (still) matter. *Personality and Social Psychology Bulletin, 38,* 209–219. http://dx.doi.org/10.1177/0146167211432763

Buckingham, J. T., & Alicke, M. D. (2002). The influence of individual versus aggregate social comparison and the presence of others on self-evaluations. *Journal of Personality and Social Psychology, 83,* 1117–1130. http://dx.doi.org/10.1037/0022-3514.83.5.1117

Buunk, B. P., Schaufeli, W. B., & Ybema, J. F. (1994). Burnout, uncertainty, and the desire for social comparison among nurses. *Journal of Applied Social Psychology, 24,* 1701–1718. http://dx.doi.org/10.1111/j.1559-1816.1994.tb01570.x

Chambers, J. R., & Windschitl, P. D. (2004). Biases in social comparative judgments: The role of nonmotivated factors in above-average and comparative-optimism effects. *Psychological Bulletin, 130,* 813–838. http://dx.doi.org/10.1037/0033-2909.130.5.813

Chen, Y., Hong, C., Chu, H. (2015). Galaxy plot and a multivariate method for correcting publication bias in multivariate meta-analysis (in preparation).

Coburn, K. M., & Vevea, J. L. (2017). Weightr: Estimating weight-function models for publication bias [Computer software]. Retrieved from https://cran.r-project.org/web/packages/weightr/index.html

Collins, R. L. (1996). For better or worse: The impact of upward social comparisons on self-evaluations. *Psychological Bulletin, 119,* 51–69. http://dx.doi.org/10.1037/0033-2909.119.1.51

Collins, R. L. (2000). Among the better ones: Upward assimilation in social comparison. In J. Suls & L. Wheeler (Eds.), *Handbook of social comparison: Theory and research* (pp. 141–158). New York, NY: Kluwer Academic/Plenum Press Publishers. http://dx.doi.org/10.1007/978-1-4615-4237-7_9

Darley, J. M., & Aronson, E. (1966). Self-evaluation vs. direct anxiety reduction as determinant of the fear-affiliation relationship. *Journal of Experimental Social Psychology, 1*(Suppl. 1), 66–79. http://dx.doi.org/10.1016/0022-1031(66)90067-9

Davis, J. A. (1966). The campus as a frog pond: An application of the theory of relative deprivation to career decisions of college men. *American Journal of Sociology, 72,* 17–31. http://dx.doi.org/10.1086/224257

De Botton, A. (2004). *Status anxiety*. New York, NY: Pantheon Books.

Dijksterhuis, A., Spears, R., Postmes, T., Stapel, D. A., Koomen, W., van Knippenberg, A., & Scheepers, D. (1998). Seeing one thing and doing another: Contrast effects in automatic behavior. *Journal of Personality and Social Psychology, 75,* 862–871. http://dx.doi.org/10.1037/0022-3514.75.4.862

Dijksterhuis, A., & van Knippenberg, A. (1998a). *The regulation of unconscious and unintentional behavior*. Unpublished manuscript, Department of Psychology, University of Nijmegen, Nijmegen, the Netherlands.

Dijksterhuis, A., & van Knippenberg, A. (1998b). The relation between perception and behavior, or how to win a game of trivial pursuit. *Journal of Personality and Social Psychology, 74,* 865–877. http://dx.doi.org/10.1037/0022-3514.74.4.865

Dillard, J. P., & Shen, L. (2005). On the nature of reactance and its role in persuasive health communication. *Communication Monographs, 72,* 144–168. http://dx.doi.org/10.1080/03637750500111815

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic traits definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57,* 1082–1090. http://dx.doi.org/10.1037/0022-3514.57.6.1082

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56,* 455–463. http://dx.doi.org/10.1111/j.0006-341X.2000.00455.x

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315,* 629–634. http://dx.doi.org/10.1136/bmj.315.7109.629

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7,* 117–140.

Gervais, W. (2015, June 25). *Putting PET-PEESE to the test*. [Blog post]. Retrieved from http://willgervais.com/blog/2015/6/25/putting-pet-peese-to-the-test-1

Gibbons, F. X., & Buunk, B. P. (1999). Individual differences in social comparison: Development of a scale of social comparison orientation. *Journal of Personality and Social Psychology, 76,* 129–142. http://dx.doi.org/10.1037/0022-3514.76.1.129

Gibbons, F. X., & Gerrard, M. (1991). Downward comparison and coping with threat. In J. Suls & T. A. Wills (Eds.), *Social comparison: Contemporary theory and research* (pp. 317–345). Hillsdale, NJ: Lawrence Erlbaum.

Gleser, L. J., & Olkin, I. (2009). Stochastically dependent effect sizes. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 357–376). New York, NY: Russell Sage Foundation.

Grabe, S., Ward, L. M., & Hyde, J. S. (2008). The role of the media in body image concerns among women: A meta-analysis of experimental and correlational studies. *Psychological Bulletin, 134,* 460–476. http://dx.doi.org/10.1037/0033-2909.134.3.460

Gruder, C. L. (1971). Determinants of social comparison choices. *Journal of Experimental Social Psychology, 7,* 473–489. http://dx.doi.org/10.1016/0022-1031(71)90010-2

Hakmiller, K. L. (1966). Threat as a determinant of downward comparison. *Journal of Experimental Social Psychology, 1*(Suppl. 1), 32–39. http://dx.doi.org/10.1016/0022-1031(66)90063-1

Heatherton, T. F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social Psychology, 60,* 895–910. http://dx.doi.org/10.1037/0022-3514.60.6.895

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1,* 39–65. http://dx.doi.org/10.1002/jrsm.5

Heider, F. (1958). *The psychology of interpersonal relations*. New York, NY: Wiley. http://dx.doi.org/10.1037/10628-000

Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions* (Version 5.1.0). Retrieved from www.handbook.cochrane.org

Hilgard, J. (2015). *Putting PET-PEESE to the test, Pt 1A* [Blog post]. Retrieved from http://crystalprisonzone.blogspot.com/2015/06/putting-pet-peese-to-test-part-1a.html

Huguet, P., Dumas, F., Marsh, H., Wheeler, L., Seaton, M., Nezlek, J., . . . Régner, I. (2009). Clarifying the role of social comparison in the big-fish-little-pond effect (BFLPE): An integrative study. *Journal of Personality and Social Psychology, 97,* 156–170. http://dx.doi.org/10.1037/a0015558

Hunter, J. P., Saratzis, A., Sutton, A. J., Boucher, R. H., Sayers, R. D., & Bown, M. J. (2014). In meta-analyses of proportion studies, funnel plots were found to be an inaccurate method of assessing publication bias. *Journal of Clinical Epidemiology, 67,* 897–903. http://dx.doi.org/10.1016/j.jclinepi.2014.03.003

Inzlicht, M., Gervais, W., & Berkman, E. (2015). *Bias-correction techniques alone cannot determine whether ego depletion is different from zero: Commentary on Carter, Kofler, Forster, & McCullough, 2015*. Retrieved from http://ssrn.com/abstract=2659409 or http://dx.doi.org/10.2139/ssrn.2659409

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation and information in hypothesis testing. *Psychological Review, 94,* 211–228. http://dx.doi.org/10.1037/0033-295X.94.2.211

Krueger, J., & Clement, R. W. (1994). The truly false consensus effect: An ineradicable and egocentric bias in social perception. *Journal of Personality and Social Psychology, 67,* 596–610. http://dx.doi.org/10.1037/0022-3514.67.4.596

Kühnen, U., & Haberstroh, S. (2004). Self-construal activation and focus of comparison as determinants of assimilation and contrast in social comparisons. *Current Psychology of Cognition, 22,* 289–310.

Manis, M., Biernat, M., & Nelson, T. F. (1991). Comparison and expectancy processes in human judgment. *Journal of Personality and Social Psychology, 61,* 203–211. http://dx.doi.org/10.1037/0022-3514.61.2.203

Marsh, H. W., & Parker, J. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology, 47,* 213–231. http://dx.doi.org/10.1037/0022-3514.47.1.213

Martin, L. L. (1986). Set/reset: Use and disuse of concepts in impression formation. *Journal of Personality and Social Psychology, 51,* 493–504. http://dx.doi.org/10.1037/0022-3514.51.3.493

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science, 11,* 730–749. http://dx.doi.org/10.1177/1745691616662243

Molleman, E., Pruyn, J., & van Knippenberg, A. (1986). Social comparison processes among cancer patients. *British Journal of Social Psychology, 25,* 1–13. http://dx.doi.org/10.1111/j.2044-8309.1986.tb00695.x

Morse, S., & Gergen, K. J. (1970). Social comparison, self-consistency, and the concept of self. *Journal of Personality and Social Psychology, 16,* 148–156. http://dx.doi.org/10.1037/h0029862

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med, 7,* e1000097. http://dx.doi.org/10.1371/journal.pmed.1000097

Mussweiler, T. (2001). Seek and ye shall find: Antecedents of assimilation and contrast in social comparison. *European Journal of Social Psychology, 31,* 499–509. http://dx.doi.org/10.1002/ejsp.75

Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review, 110,* 472–489. http://dx.doi.org/10.1037/0033-295X.110.3.472

Mussweiler, T., & Strack, F. (2000). Consequences of social comparison: Selective accessibility, assimilation, and contrast. In J. Suls & L. Wheeler (Eds.), *Handbook of social comparison: Theory and research* (pp. 253–270). New York, NY: Kluwer Academic/Plenum Press Publishers. http://dx.doi.org/10.1007/978-1-4615-4237-7_13

Myers, T. A., & Crowther, J. H. (2009). Social comparison as a predictor of body dissatisfaction: A meta-analytic review. *Health Psychology, 118,* 683–698. http://dx.doi.org/10.1037/a0016763

Norton, M. I. (2013). All ranks Are local: Why humans are both (painfully) aware and (surprisingly) unaware of their lot in life. *Psychological Inquiry, 24,* 124–125. http://dx.doi.org/10.1080/1047840X.2013.794689

Orwin, R. G. (1983). A fail-safe *N* for effect size in meta-analysis. *Journal of Educational Statistics, 8,* 157–159. http://dx.doi.org/10.2307/1164923

Petty, R., & Wegener, D. (1993). Flexible correction processes in social judgment. *Journal of Experimental Social Psychology, 29,* 137–165. http://dx.doi.org/10.1006/jesp.1993.1007

Pleban, R., & Tesser, A. (1981). The effects of relevance and quality of another's performance on interpersonal closeness. *Social Psychology Quarterly, 44,* 278–285. http://dx.doi.org/10.2307/3033841

Pustejovsky, J. (2016). clubSandwich [Computer software]. Retrieved from https://cran.r-project.org/web/packages/clubSandwich/

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press. http://dx.doi.org/10.1515/9781400876136

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 85,* 638–641. http://dx.doi.org/10.1037/0033-2909.86.3.638

Schachter, S. (1959). *The psychology of affiliation*. Redwood City, CA: Stanford University Press.

Schwarz, N., & Bless, H. (1992). Constructing reality and its alternatives: An inclusion/exclusion model of assimilation and contrast effects in social judgment. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 217–245). Hillsdale, NJ: Lawrence Erlbaum.

Sedikides, C., Gaertner, L., & Vevea, J. L. (2005). Pancultural self-enhancement reloaded: a meta-analytic reply to Heine (2005). *Journal of Personality and Social Psychology, 89,* 539–551. http://dx.doi.org/10.1037/0022-3514.89.4.539

Shanks, D. R., Newell, B. R., Lee, E. H., Balakrishnan, D., Ekelund, L., Cenac, Z., . . . Moore, C. (2013). Priming intelligent behavior: An elusive phenomenon. *PLoS ONE, 8,* e56515. http://dx.doi.org/10.1371/journal.pone.0056515

Smith, E. R., & Branscombe, N. R. (1987). Procedurally-mediated social inferences: The case of category accessibility effects. *Journal of Experimental Social Psychology, 23,* 361–382. http://dx.doi.org/10.1016/0022-1031(87)90036-9

Srull, T. K., & Gaelick, L. (1983). General principles and individual differences in the self as a habitual reference point: An examination of self-other judgments of similarity. *Social Cognition, 2,* 108–121. http://dx.doi.org/10.1521/soco.1983.2.2.108

Stanley, T. (2017). Limitations of PET-PEESE and other meta-analysis methods. *Social Psychological & Personality Science*. Advance online publication. http://dx.doi.org/10.1177/1948550617693062

Suls, J., Martin, R., & Wheeler, L. (2002). Social comparison: When, with whom and with what effect? *Current Directions in Psychological Science, 11,* 159–163. http://dx.doi.org/10.1111/1467-8721.00191

Suls, J., & Wheeler, L. (2017). On the trail of social comparison. In K. Williams, S. Harkins, & J. Berger (Eds.), *Oxford handbook of social* *influence*. New York, NY: Oxford University Press. http://dx.doi.org/10.1093/oxfordhb/9780199859870.013.13

Tennen, H., & Affleck, G. (1997). Social comparison as a coping process: A critical review and application to chronic pain disorders. In B. P. Buunk & F. X. Gibbons (Eds.), *Health, coping, and well-being: Perspectives from social comparison theory* (pp. 263–298). Hillsdale, NJ, USA: Erlbaum.

Tennen, H., McKee, T. E., & Affleck, G. (2000). Social comparison processes in health and illness. In J., Suls & L., Wheeler (Eds.), *Handbook of social comparison: Theory and research* (pp. 443–486). New York, NY: Kluwer Academic/Plenum Publishers.

Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine, 22,* 2113–2126. http://dx.doi.org/10.1002/sim.1461

Tesser, A. (1988). Toward a self-evaluation maintenance model of social behavior. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 21, pp. 181–227). New York, NY: Academic Press. http://dx.doi.org/10.1016/S0065-2601(08)60227-0

Thornton, D. A., & Arrowood, A. J. (1966). Self-evaluation, self-enhancement and the locus of social comparison. *Journal of Experimental Social Psychology, 1*(Supp. 1), 40–48. http://dx.doi.org/10.1016/0022-1031(66)90064-3

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352. http://dx.doi.org/10.1037/0033-295X.84.4.327

Upshaw, H. S. (1978). Social influences on attitudes and on anchoring of congeneric attitude scales. *Journal of Experimental Social Psychology, 14,* 327–339. http://dx.doi.org/10.1016/0022-1031(78)90029-X

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods, 45,* 576–594. http://dx.doi.org/10.3758/s13428-012-0261-6

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods, 10,* 428–443. http://dx.doi.org/10.1037/1082-989X.10.4.428

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36,* 1–48. http://www.jstatsoft.org/v36/i03/

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54,* 1063–1070. http://dx.doi.org/10.1037/0022-3514.54.6.1063

Wheeler, L. (1966). Motivation as a determinant of upward comparison. *Journal of Experimental Social Psychology, 1*(Supp. 1), 27–31. http://dx.doi.org/10.1016/0022-1031(66)90062-X

Wheeler, L., & Miyake, K. (1992). Social comparison in everyday life. *Journal of Personality and Social Psychology, 62,* 760–773. http://dx.doi.org/10.1037/0022-3514.62.5.760

Wheeler, L., & Reis, H. T. (1991). Self-recording of everyday life events: Origins, types, and uses. *Journal of Personality, 59,* 339–354.

Wheeler, L., & Suls, J. (2007). Assimilation in social comparison: Can we agree on what it is? *Revue Internationale de Psychologie Sociale, 20,* 31–51.

Wills, T. A. (1981). Downward comparison principles in social psychology. *Psychological Bulletin, 90,* 245–271. http://dx.doi.org/10.1037/0033-2909.90.2.245

Wood, J. V. (1989). Theory and research concerning social comparisons of personal attributes. *Psychological Bulletin, 106,* 231–248. http://dx.doi.org/10.1037/0033-2909.106.2.231

Wood, J. (1996). What is social comparison and how should we study it? *Personality and Social Psychology Bulletin, 22,* 520–537. http://dx.doi.org/10.1177/0146167296225009

Wood, J. V., Taylor, S. E., & Lichtman, R. R. (1985). Social comparison in adjustment to breast cancer. *Journal of Personality and Social Psychology, 49,* 1169–1183. http://dx.doi.org/10.1037/0022-3514.49.5.1169

Zell, E., & Alicke, M. D. (2010). The local dominance effect in self-evaluation: Evidence and explanations. *Personality and Social Psychology Review, 14,* 368–384. http://dx.doi.org/10.1177/1088868310366144

Zell, E., & Alicke, M. D. (2013). Local dominance in health risk perception. *Psychology & Health, 28,* 469–476. http://dx.doi.org/10.1080/08870446.2012.742529

Zuckerman, M. (1960). The development of an affect adjective check list for the measurement of anxiety. *Journal of Consulting Psychology, 24,* 457–462. http://dx.doi.org/10.1037/h0042713

Received November 5, 2015
Revision received September 8, 2017
Accepted September 11, 2017 ■