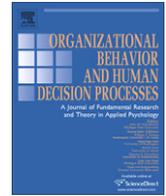




Contents lists available at ScienceDirect

Organizational Behavior and Human Decision Processes

journal homepage: www.elsevier.com/locate/obhdp

Testing four explanations for the better/worse-than-average effect: Single- and multi-item entities as comparison targets and referents

Jerry Suls^{a,*}, John Chambers^b, Zlatan Krizan^c, Chad R. Mortensen^d, Bryan Koestner^e, Kathryn Bruchmann^e

^a Department of Psychology, University of Iowa, Iowa City, IA 52242-1407, United States

^b University of Florida, Gainesville, FL, United States

^c Iowa State University, Ames, IA, United States

^d Arizona State University, Tempe, AZ, United States

^e University of Iowa, Iowa City, IA, United States

ARTICLE INFO

Article history:

Received 22 April 2008

Accepted 25 March 2010

Available online 21 May 2010

Accepted by Maurice Schweitzer

Keywords:

Better-than-average effect

Worse-than-average effect

Comparative bias

Social comparison

Focalism

ABSTRACT

In six experiments, we tested four explanations for the better/worse-than-average effect (B/WTA) by manipulating the number of items comprising the target or referent of direct comparison. A single-item target tended to be rated more extremely than a single-item or a multi-item referent (Experiments 1–3). No B/WTA was obtained, however, when a multi-item target was compared with either a single- or multi-item referent (Experiments 4 and 5). A bias favoring a multi-item target was found only if cohesiveness among the items was increased through instructions (Experiment 6). The Unique-Attributes Hypothesis generally provided the best explanation the findings; the focalism explanation also demonstrated some empirical viability. The results suggest that important preferential decision-making outcomes can be affected by both the number of items and whether items are strategically manipulated to serve as targets or referents of comparison.

© 2010 Elsevier Inc. All rights reserved.

Introduction

The *better-than-average effect* is a widely replicated finding whereby people rate themselves as better than their peers on a variety of dimensions (Alicke, 1985; Codol, 1975; Klein & Weinstein, 1997; Taylor & Brown, 1988). Although this effect was originally conceptualized as an expression of a self-enhancement motive, people also rate casual acquaintances as better than average (Klar, Medding, & Sarel, 1996). For example, a randomly selected peer from a favored group is rated as better than other members of the group (Klar & Giladi, 1997). Even a randomly selected item from a set of positive category exemplars is rated as better than the rest (e.g., appealing songs, soaps, sofas, vacation destinations or foods; see Chambers and Suls (2007), Giladi and Klar (2002), Windschitl, Conybeare, and Krizan (2008).

To complicate the picture, *worse-than-average* effects, which clearly are inconsistent with self-enhancement, have also been found. For example, college students rate themselves *worsethan* the average student on difficult tasks such as juggling and computer programming (Kruger, 1999; see also Moore & Kim, 2003; Windschitl, Kruger, & Simms, 2003). People also rate a randomly selected exemplar from a negative category (e.g., unappealing songs or irri-

tating people) more negatively than the rest (Giladi & Klar, 2002; Moore & Kim, 2003)—situations where the self is not involved.

The similar pattern for positive and negative category members points to the existence of a general comparative bias, whereby any randomly selected target from a positive or negative category is rated more extremely on the (positive or negative) evaluative dimension when compared directly with others from the same category, considered as a whole. Giladi and Klar (2002) refer to this general phenomenon as the Non-Selective Superiority/Inferiority Bias (NSS/IB), but we will use the more generic, B/WTA effect. Our reason for this is that NSS/IB specifically involves comparison of a single-item target vs. a multi-item referent, which is the most commonly researched comparison scenario in this field of study. The experiments reported below also involve other comparison scenarios—when the referent is a single-item and the target consists of multiple-items or when both the target and the referent have single- or multiple-items.

The B/WTA has implications for a broad range of situations where comparative judgments are made, such as employment settings, self-evaluation and political decision-making (e.g., Johnson, 2004; Malmendier & Tate, 2005). For example, if a pool of job applicants all possess good qualifications, any one of them who happens to become the target of comparison may be judged better than the rest. The consequence is that the most qualified candidate may not receive the appropriate advantage, and may ultimately not be selected. Consequently, a full understanding of the processes underlying this comparative bias is needed.

* Corresponding author.

E-mail address: jerry-suls@uiowa.edu (J. Suls).

Several explanations have been proposed for the B/WTA (see Chambers & Windschitl, 2004; Moore, 2007). One emphasizes the role of *focalism*, which refers to the tendency to pay greater attention to information in the foreground or to the salient hypothesis than to information in the background or to alternative hypotheses (e.g., Chambers & Windschitl, 2004; Moore & Kim, 2003; Skade & Kahneman, 1998). Applied to a direct comparison (e.g., *Do you like French fries more than other fast foods?*), more weight should be given to the target of the comparison (i.e., fries) than the referent (i.e., other fast foods) because, according to focalism, more information should be recruited about the target's positive features than about the referent's positive features. Consistent with this reasoning, an item is rated more positively (negatively) if it serves as the target than as the referent (e.g., Chambers & Suls, 2007; Eiser, Pahl, & Prins, 2001; Hoorens, 1995; Krizan & Suls, 2008). Also, absolute ratings of the target (i.e., ratings in response to *"How much do you like French fries?"*) tend to be more strongly associated with comparative ratings than are absolute ratings of the referent (*"How much do you like other fast foods?"*) (Eiser et al., 2001; Klar, 2002; Klar & Giladi, 1999; Kruger, 1999; Kruger & Burrus, 2004; Moore & Kim, 2003). This pattern suggests that information about the target receives greater weight in the comparative judgment than information about the referent (a response pattern that may be logically defensible under some circumstances, see e.g., Chambers and Windschitl (2004), Kruger, Windschitl, Burrus, Fessel, and Chambers (2008), Moore (2007)). Moore and Kim (2003) refer to this focus on the target as a kind of "comparison myopia."

A second, non-mutually exclusive, explanation (Chambers & Windschitl, 2004; Moore, 2007) emphasizes the diffuse nature of the referent. The *group diffuseness* idea is that an entity with multiple-items (i.e., "the average student", "all soaps in this set") is more difficult to thoughtfully and accurately evaluate than an entity with only a single-item (e.g., "you", "your friend", "this soap"). Consequently, when a single-item entity is compared with a multi-item entity, the former will be rated more extremely, regardless of whether it is the target or referent of comparison.

Consistent with this idea, some studies have found the B/WTA to be reduced or eliminated when the self is compared to a specific individual rather than a group (Alicke, Klotz, Breitenbecher, Yurak, & Vredenburg, 1995; Klar et al., 1996; Perloff & Fetzler, 1986). Other investigators (Moore & Kim, 2003; Regan, Snyder, & Kassir, 1995; Suls, Lemos, & Stewart, 2002; Windschitl et al., 2003), however, find the B/TWA even when participants compared themselves with a specific individual. The reasons for the inconsistency among results are unclear because experiments employed different kinds of exemplars (e.g., the self, acquaintances, open-ended categories, such as the "average college student," objects).

Giladi and Klar (2002) propose the Local Comparisons-General Standards (LOGE) model for the B/TWA. According to LOGE, when people compare a single-item with other items from the same (positive or negative) group, they "... fail to use the normatively appropriate local (group) standard and are infelicitously affected by a more general standard" (p. 538). For example, when comparing how irritating one person is relative to other irritating people, the rater relies on a hybrid standard, composed not only of the other irritating people but also "other people in general". The addition of people in general creates a standard that is less irritating overall so the target is rated more extremely. LOGE posits that the encroachment of a global standard affects direct comparison judgments. Consistent with this explanation, the bias is not manifested when multi-item entities are heterogeneous rather than homogeneous superior (positive) or inferior (negative) (Giladi & Klar, 2002, Exp. 5). This follows from LOGE because in a heterogeneous group some members are above the general standard, some below the standard and some about the standard. As a consequence, the hybrid standard would be closer to the local standard

in heterogeneous groups; the single-item should therefore not be rated more extremely.

A fourth explanation is the Unique-Attributes Hypothesis (U-AH; Chambers, *in press*), which is based on the idea that each item in the comparison will share attributes (by definition) with other items, but also will probably differ on at least one attribute. For U-AH, when an item is compared with others, its unique attribute receives more weight than other relevant attributes. "In other words, the comparison may be made primarily on the dimension or attribute that most distinguishes the target from other comparison members ... [thus] ... almost any member of a positively valenced group is judged superior to the rest of its group (because its unique positive attribute becomes the comparison standard) and almost any member of a negatively valenced group is judged inferior to the rest of its group (because its unique negative attributes becomes the comparison standard)," (p. 2). Consistent with the U-A hypothesis, thought protocols and implicit measures of accessibility showed when that targets (i.e., popular or unpopular vacation destinations) were compared with each other, their unique attributes were used as the comparison standard (Chambers, *in press*).

The present experiments were conducted because these four explanations for the B/TWA can be empirically differentiated as a function of whether a single-item or a multi-item entity serves in the role of comparison target or referent. The different predictions derived from the four approaches are described below (also see Table 1).

According to focalism, a target should receive more weight in the comparison (and judged more extremely than the referent) because of selective recruitment of information favoring the target. Thus, a target should be judged more extremely, regardless of whether the target or referent consists of a single-item or multiple-items.

For the group diffuseness explanation, a single-item entity should have a judgmental advantage as a target or referent because of the cognitive demands typically associated with forming an overall impression of multiple-items. Thus, a single entity should be rated more extremely whether it serves as target or referent. In fact, if a multi-item target is compared to a single-item referent, a reverse bias (favoring the single-item referent) should be obtained because the single-item should have the information processing advantage.

Whether the target is a single-item or multiple-items, LOGE posits that the local and the general standard should be the same. This means the B/TWA should be exhibited regardless of whether a target or referent is a single- or multi-item entity.

The Unique-Attributes Hypothesis posits that more weight is given to the unique attribute of the target (than referent) of the comparison, and that recruiting a unique attribute is easier for a single-item than multi-item entity. Therefore, a single-item target should be rated more extremely than either a single- or multi-item referent (particularly with the latter). Because generating a unique attribute for a multi-item entity is difficult, a multi-item target should be rated no more extremely than a multi-item referent, and may actually be rated slightly less extreme than a single-item referent (i.e., a small reversal in the direction and magnitude of bias).

Six experiments were conducted to test these differential predictions (see Table 1 for summary of predictions). In the initial studies, we systematically manipulated target and referent size. To preview the results, we found a multi-item entity consistently had a judgmental disadvantage. In a final experiment, two manipulations hypothesized to remove the disadvantage were tested.

Experiment 1

The initial experiment was conducted to assess the effects of the number of items comprising the referent; the comparison target was always a single-item. Participants were asked to generate a list

Table 1
Predictions about type of target and referent.

Type of target	Type of referent	Predicts a non-selective bias favoring the target?			
		Unique attributes	Focalism	Group diffuseness	Local-general (LOGE)
Single	Single	Yes (small bias)	Yes	No	Yes
Single	Group	Yes	Yes	Yes	Yes
Group	Single	No (small reverse bias)	Yes	No (large reverse bias)	Yes
Group	Group	No	Yes	No	Yes

of seven favorite foods and subsequently rated a randomly selected food from the list vs. the six others, two others or only one other. The last case, where participants rated a single target (randomly selected) vs. a single referent, provided a strong test of the group diffuseness account. According to this account, if people give a single target more judgmental weight solely because a multi-item referent is diffuse and therefore difficult to consider in aggregate, the bias should be eliminated when the referent consists only of a single-item. Focalism, LOGE, and U-AH predict the B/TWA will be exhibited even if both the target and referent are single-items (according to U-AH, because more weight is given to the target than referent's unique attribute).

Method

Participants and design

One hundred twenty-four undergraduate students (in small groups ranging from one to eight) from an introductory psychology course participated and were randomly assigned to the 1 vs. 6, 1 vs. 2 or 1 vs. 1 conditions.

Materials and procedure

Upon arrival at the laboratory, participants were informed that the study concerned people's preferences for various foods. Each participant was given a folder containing a manila envelope (two envelopes in the 1 vs. 2 and 1 vs. 1 conditions), seven slips of paper and an instruction sheet. In all conditions, the instruction sheet requested that participants think of seven of their favorite foods and write each on a slip of paper. Next, participants were instructed to place all slips of paper in the envelope (marked 'A'). This two-stage procedure was used so that all participants initially generated the same number of stimuli. This procedure was used in all of the subsequent experiments.

Participants were then asked to shake the envelope (marked 'A') well, so that all the slips were shuffled. In the 1 vs. 2 and 1 vs. 1 conditions, participants were then instructed to pull out, without looking, either four or five slips, respectively, and place them in the other envelope (marked 'B') and put it aside. Then, participants in all conditions removed the slips from the remaining envelope ("A") and reviewed them. Following this, participants returned the slips, shook the envelope again, and without looking, drew one slip of paper out. This ensured that participants were aware that the food listed on that slip had been chosen at random. Finally, participants were instructed to respond to the three questions of interest printed on a separate sheet of paper.

Measures

The first of three questions involved a target-to-referent comparison and read, "Compared to the other food[s] that you listed (remaining in envelope 'A'), how would you rate the food that you just drew from the envelope?" Participants made ratings on an 11-point Likert-type scale ranging from 0 = dislike it much more than the other food(s) to 5 = like it as much as the other food(s) to 10 = like it much more than the other food(s). Participants in the 1 vs. 6 condition rated the target compared to the other six stimuli, those in the 1 vs. 2 condition rated the target compared to the

remaining two stimuli and those in the 1 vs. 1 condition rated the target compared to the single remaining stimulus.

The next question concerned the absolute evaluation of the target: *If you had to give a score to the food that you just drew from the envelope, how would you score it?* An 11-point Likert-type scale was provided, with the verbal labels, *extremely dislike it* (0), *neither like it nor dislike it* (5), and *extremely like it* (11). Finally, participants provided an absolute evaluation of the referent group of foods (or a single food) they had generated. For example, in the 1 vs. 2 foods condition, the absolute question was phrased, *"If you had to give a score to the 2 foods remaining in the envelope 'A' as a whole, how would you score the group of two?"* Response scales were similar to the ones used for the absolute ratings of the target. Finally, participants were debriefed and dismissed.

Results and discussion

Manipulation checks

Absolute evaluations of the randomly selected target food ($M = 8.5$; $SD = 1.2$) and the group ($M = 8.5$; $SD = 1.2$) were positive (indicated by a significant difference from the midpoint, 5, p 's < .001). Thus, participants generated favorite foods, as instructed. The absolute ratings of the target and referent did not differ, $t(123) = -.05$, *ns*. Also, neither the absolute ratings of the target or the referent differed as a function of number of items comprising the referent (F s < 1, *ns*), which may be seen as inconsistent with the group diffuseness account (because a multi-item entity is presumably more difficult to evaluate, and hence should be evaluated less extremely). In sum, considered separately, the target and the referent were perceived as equivalent in value.

Comparative judgments and relationships with absolute ratings

Participants' responses to the comparative question were recoded to a -5 to +5 scale for ease of interpretation before submitting them to statistical analyses. Matched t -tests indicated comparative biases in the 1 vs. 6 ($M = 1.7$, $SD = 1.8$, $t(40) = 6.1$, $p < .001$), 1 vs. 2 ($M = 1.9$, $SD = 2.1$, $t(42) = 5.4$, $p < .001$), and 1 vs. 1 conditions ($M = 1.8$, $SD = 2.1$, $t(39) = 5.4$, $p < .001$). In each condition, the randomly selected target was rated more positively than the referent. The number of items comprising the referent had no effect on the magnitude of this bias $F(2) = .13$, *ns*; a single-item target was rated more positively whether it was compared with a single-item or multi-item referent (comprised of six items) from the same category.

Zero-order correlations between absolute and comparative ratings are presented in Table 2. The pattern of associations did not vary between experimental conditions, except that the correlation between comparison rating and absolute rating of the referent was non-significant in the 1 vs. 1 condition, $r(40) = -.08$, $p > .10$. Even in this condition, however, the absolute rating of the target was strongly related to the comparative rating. According to focalism, absolute liking for the target food should be a consistently stronger predictor of comparative ratings than absolute liking of the referent, even when the referent is a single-item entity (1 vs. 1 condition). LOGE and UA-H are mute on the issue of the relative weighting given to each entity (target or referent), the former because it posits a combined local plus global standard, and the latter

Table 2

Zero-order correlations between judgments as a function of comparison set size in Experiment 1 (total $N = 124$).

	Comparative	Target	Referent
<i>1 vs. 6 (N = 41)</i>			
Comparative	–		
Target	.63**	–	
Referent	–.32*	.09	–
<i>1 vs. 2 (N = 43)</i>			
Comparative	–		
Target	.67**	–	
Referent	–.33**	–.14	–
<i>1 vs. 1 (N = 40)</i>			
Comparative	–		
Target	.64**	–	
Referent	–.08	.27	–

* $p < .05$.

** $p < .01$.

because it posits that both entities are evaluated according to the target’s unique attribute. Since the target was rated more extremely than a referent that consisted of a single-item, the group diffuseness explanation received no support.

Experiment 2

This second experiment was a replication and extension testing the effects of number of number of items comprising the referent on the B/WTA, but involved judgments of social stimuli (i.e., other people) and both positive and negative categories.

Method

Overview

Participants were asked to generate a list of six pleasant or six irritating acquaintances, then one acquaintance was randomly selected from the list and compared to: (a) the remaining five acquaintances, (b) a subset of two acquaintances, or (c) one acquaintance from the same list.

Participants and design

One-hundred and seventy-six undergraduate students (in small groups from one to eight) participated and were randomly assigned to one cell of a 2 (acquaintance type: pleasant or irritating) \times 3 (referent size: 6, 2, or 1) factorial design.

Materials and procedure

Upon arrival at the laboratory, participants were informed that the study concerned people’s judgments of others. Each participant was given a folder containing one manila envelope (two envelopes in the 1 vs. 2 and 1 vs.1 conditions), six slips of paper, and an instruction sheet. The instruction sheet requested that participants write the names of six pleasant or irritating acquaintances on the slips of paper. Next, the instructions indicated that participants should place all slips in the envelope (marked ‘A’) that had been provided. The rest of the procedure was the same as in Experiment 2, concluding with participants responding to the three questions of interest.

The first question, which involved a target-to-referent comparison, read, “Compared to the other people that you listed, how would you rate the person that you just drew from the envelope?” Participants provided responses on an 11-point Likert-type scale ranging from 0 to 10. Scale anchors were similar to those in the previous experiment, although the anchors were adjusted to correspond to the number of people participants were comparing in specific conditions and to the appropriate category (pleasant vs. irritating). The question for the absolute evaluation of the target and its scale

anchors were similar to that used Experiment 1, with the upper end of the scale always reflecting more of the quality under consideration (e.g., regarding irritating acquaintances, endorsing higher scale values for the target meant judging the acquaintance to be more irritating). Finally, participants provided an absolute evaluation of the referent (5, 2 or 1) they had generated. For example, in the 1 vs. 5 condition, the question was “If you had to give a score to the group of people as a whole, how would you score it?” The response scales were similar to those used in the prior study with appropriate adjustments to both the question and anchors, as required by the condition. Again, the upper end of the scale reflected more of the relevant quality (i.e., more pleasant or more irritating). Finally, participants were debriefed and dismissed.

Results and discussion

Manipulation checks

Across all of the pleasant acquaintances conditions, the absolute ratings of both the target ($M = 7.9$, $SD = 2.3$) and the referent ($M = 7.8$; $SD = 1.8$) were positive, $ps < .001$. Likewise, across all of the unpleasant acquaintances conditions, absolute evaluations of the target ($M = 7.6$; $SD = 1.8$) and referent ($M = 7.4$; $SD = 2.0$) were negative, $ps < .001$. Ratings of targets and their referents did not differ significantly (all $ts < |1.3|$), and the extremity of both sets of absolute ratings did not vary as a function of group size or acquaintance type (all $Fs < 1$). As in Experiment 1, separate ratings of the single-item targets and multi-item referents showed no significant differences, which may be seen as inconsistent with the group diffuseness account.

Comparative judgments and relationships to absolute ratings

As before, the comparative ratings were transformed to a –5 to 5 scale. Targets in the pleasant acquaintance conditions were rated more positively than the referent in the 1 vs. 5 condition ($M = 1.3$, $SD = 2.0$, $t(26) = 3.3$, $p < .01$), 1 vs. 2 condition ($M = 0.9$, $SD = 1.6$, $t(29) = 3.0$, $p < .01$), and marginally in the 1 vs. 1 condition ($M = 0.5$, $SD = 1.8$, $t(32) = 1.7$, $p = .10$).

Irritating targets were rated more negatively in the 1 vs. 5 condition ($M = 1.4$, $SD = 2.9$, $t(29) = 2.6$, $p < .05$), 1 vs. 2 ($M = 1.8$, $SD = 2.6$, $t(29) = 2.4$, $p < .05$), and 1 vs. 1 conditions ($M = 1.3$, $SD = 2.8$, $t(25) = 2.4$, $p < .05$). The magnitude of the bias did not vary as a function of number of items comprising the referent in either the pleasant, $F(2) = 1.4$, ns , or irritating acquaintances conditions, $F(2) = .04$, ns .

Table 3

Zero-order correlations between judgments as a function of acquaintance type and comparison set size in Experiment 2 ($N = 176$).

	Pleasant acquaintances			Irritating acquaintances		
	Comp	Target	Referent	Comp	Target	Referent
<i>1 vs. 5 (N = 57)</i>						
Comp	–			–		
Target	.69**	–		.67**	–	
Referent	.50**	.78**	–	–.10	.31*	–
<i>1 vs. 2 (N = 60)</i>						
Comp	–			–		
Target	.41**	–		.56**	–	
Referent	–.03	.73**	–	–.25	.45**	–
<i>1 vs. 1 (N = 59)</i>						
Comp	–			–		
Target	.54**	–		.90**	–	
Referent	.14**	.77**	–	–.62**	–.44**	–

* $p < .05$.

** $p < .10$.

** $p < .01$.

There was consistent evidence for the B/WTA, regardless of the number of items in the referent—in contradiction to the diffuse-group explanation. For the 1 vs 1 positive condition, the bias was only marginally significant (see above), which might be seen as most consistent with UA-H because, as we suggested, in such a situation the target's unique attribute may receive more weight than the referent's unique attribute.

Zero-order correlations between judgments are presented in Table 3 for all six conditions. As in the Experiment 1, the comparative ratings were more strongly associated with absolute ratings of the target than with absolute ratings of the referent. Also, as would be expected from the magnitude and direction of the comparative ratings, the absolute target and referent ratings tended to be positively and negatively correlated (respectively) with the comparative ratings. There were, however, two exceptions where the referent rating was positively correlated with the comparative rating (i.e., 1 vs. 5/pleasant and 1 vs. 1/pleasant conditions). Klar and Giladi (1999) have also reported positive correlations between absolute group ratings and comparative ratings.

Experiment 3

The two previous experiments found some evidence for a B/WTA bias even when the referent was comprised of a single-item, a result inconsistent with group diffuseness. A possible limitation of the preceding studies is that the main dependent variable involved scale ratings using subjective verbal labels. This is the most common way the B/TWA has been studied, but critics have observed that such scales are ambiguous and subject to personal interpretation, which may contribute to the emergence of B/WTA (Burson, Larrick, & Klayman, 2006; Moore, 2007). To address this possibility, Experiment 3 was a replication using a more objective judgmental metric (following Moore and Kim (2003)).

Method

Overview

The experiment involved a between-subjects design with two conditions. All participants were instructed to generate a list of six of their favorite films, but in one condition, they compared one randomly selected film from the list with the remaining five films. In the other condition, they compared one randomly selected film vs. another randomly selected film. The comparison judgment involved a dependent variable that avoided the use of ambiguous verbal labels.

Participants

Thirty-five undergraduate students (in groups of one to eight) participated in this experiment.

Materials and procedures

All participants were instructed to think of their six favorite films and write the name of each on a separate slip of paper. Each participant was also given a manila envelope (or two envelopes in the 1 vs. 1 condition). Next, the instructions asked participants to place all six slips in the envelope (marked 'A'). They were then asked to shake the envelope (marked 'A') so that all the slips were shuffled. In the 1 vs. 5 condition, participants were asked to pull out one slip and place it on one side of their desk. Then they were asked to remove the remaining five slips and place them on the other side of the desk. The instructions for the 1 vs. 1 condition were different: participants were first asked to pull out, without looking, four slips from "A", and place them in another envelope (marked 'B') and to put it aside. Then, the participant was asked to remove one slip from "A" and place on one side of the desk and then remove the remaining slip and place it

on the other side of the desk (left or right side of desk was counter-balanced across subjects).

Participants were then asked to read a second page of instructions, which included the dependent measure. The instructions described the Internet Movie Data Base (IMDb), a website that collects information about films, including the opinions of other viewers in which films are rated on a scale from 0 to 10. In the 1 vs. 5 film condition, participants were asked to: "Make your best guess about how many points (higher or lower) is the average rating of the one film compared to the average of the ratings of the five films." Then participants were presented with a 11-point scale with -5 points lower than the average of the other five films as one anchor, same points (in the middle) and $+5$ points higher than the average of the other films as the other anchor.

In the 1 vs. 1 film condition, participants were instructed to "Make your best guess about many points (higher or lower) is the average rating of the film on the one side compared to the average rating of the film on the other side." The scale was the same as in the 1 vs. 5 condition, except that one anchor was labeled " -5 points lower than the average rating of the other film" and the other anchor was " $+5$ points higher than the average rating of the other film." By using this kind of rating, each scale point represented a specific numerical value and thereby avoided subjective verbal labels. After participants made their judgments, they were debriefed and dismissed.

Results and discussion

The target film was rated more positively than the referent ($M = 1.2$, $SD = 2.3$) in the 1 vs. 5 condition, one-sample $t(18) = 2.3$, $p < .01$. Thus, the BTA effect was replicated with a more objective scale. In the 1 vs. 1 condition, the target film was also rated more positively than the referent, but only marginally ($M = 0.8$, $SD = 2.7$). The 1 vs. 1 comparison was not as robust, supporting predictions of the UA-H account (see Table 1), but the general effect found in the two previous experiments was replicated. Regardless of whether a subjective or objective scale was used, a single-item target tended to be rated more extremely whether compared to a single-item or multi-item referent. Consequently, it seems unlikely that the effects are attributable to the use of subjective rating scales¹.

¹ An anonymous reviewer had a concern about Experiment 3; instead of replacing the rating scale, we put meaningful numbers on it, leaving the problem that the scale constrains responses (i.e., since the scale only went from -5 to $+5$, it cut off more extreme responses). In response, the reviewer re-did Experiment 3 using our methods. With his/her permission, a description of the reviewer's methods and results are provided: Fifty-nine volunteers participated for a fixed \$3 payment each. Participants were instructed to write the names of six of their favorite movies on slips of paper. They were then instructed to put these slips in an envelope and shake the envelope. With their eyes closed, then drew one slip from the envelope at random and placed it on left side of the desk in front of them. They then took the other five slips and arranged them on the right side of the desk. Participants then answered the following questions: (1) make your best guess about how many points (higher or lower) is the average rating of the film on left side compared to the average rating of the set of films on the right side. (from " -5 points lower than the average rating of the set" to " $+5$ points higher than the average rating of the set"). (2) Please make an exact guess about the difference between IMDb ratings for the one movie on the left compared with the average set of the five movies on the right. This second question was followed by the clarifying follow-up: "Just to be clear, please indicate whether you think the one movie has a higher or lower IMDb rating than the set of five." The order of these two questions was counterbalanced. The second question did not constrain their responses, and so should have produced better data. But just to be sure, the reviewer also asked a subset of participants (30 people) to make individual assessments of the IMDb ratings of each of: (1) the one movie and (2) the set of five. The reviewer replicated our results exactly. On both measures, participants' comparison favoring the one movie. The ratings on the 11-point scale have the one movie 1 point up, which is significantly different from zero, $t(58) = 4.0$, $p < .001$. Estimated differences on the second question favor the one movie by 1.2 points, which is also significantly higher than zero, $t(58) = 2.93$, $p = .005$ (a result that was not exactly in line with our ideas was the indirect difference [measured by subtracting the individual rating of the 5 from the individual rating of the one] correlated highly with both and produced the same result. This indirect comparison also gives the advantage to the one movie by .72 points, $t(29) = 2.18$, $p = .04$).

Experiment 4

The results of the preceding experiments do not confirm group diffuseness predictions—a single target was rated more extremely compared to a single-item or multi-item referent. Experiment 4 was conducted to test the other predictions presented in Table 1. In Experiment 4, the target consisted of either two or five items and the referent consisted of only a single-item. According to focalism, whatever entity serves as the target should receive more judgmental weight, independent of number of items. Thus, targets consisting of two or five items should be rated more extremely than a single-item serving as the referent. In contrast, the diffuse-group idea predicts just the opposite—that a single-item referent should be rated more extremely than a multi-item target because of the judgmental disadvantage (i.e., cognitive complexity) associated with forming an impression of multi-item entities. LOGE assumes that the local and global standard should be the same, regardless of the number of items or of target/referent status so B/TWA should be found in both 2 vs. 1 and 5 vs. 1 comparisons. The U-AH approach also does not predict a B/TWA for a multi-item target because it cannot be readily conceived in terms of a unique attribute. In fact, in comparison with a single-item referent, the referent might be rated more extremely (i.e., a reverse bias). This is not expected to be a large effect, however, because we assume the unique attribute of the referent normally receives less weight than the unique attribute—if there is one—of the target (see Table 1, Row 3-Column 3).

These predictions were tested in a 2×2 between-participant design. Because the aim of the study was to determine how people compare a multi-item target with a single-item referent, the 1 vs. 1 condition was dropped from the design. Participants generated the names of six pleasant or irritating acquaintances and then compared 5 vs. 1 or 2 people vs. 1 of those acquaintances.

Overview

The design and procedures were the same as in Experiment 2 except that the target and referent were reversed in the comparison judgment question and the 1 vs. 1 condition was not included.

Method

Participants and design

One-hundred and fourteen undergraduate students (in groups from one to eight) participated and were randomly assigned to one cell of a 2 (acquaintance type: pleasant or irritating) \times 2 (comparison set size: 5 or 2) factorial design.

Materials and procedure

The procedure was the same as in Experiment 2 with the only changes involving the comparison judgment question—switching the position of the target and the referent in the comparison question. For example, the comparison question in the 2 vs. 1 condition was *Compared to the person you just drew from envelope “B,” (referring to the referent) how would you rate the other two people remaining in the envelope? (referring to the target).* Participants provided ratings on an 11-point Likert-type scale ranging from 0 to 10. Scale anchors followed a similar structure as in the previous experiment, although they were adjusted to refer to the person drawn from the envelope, e.g., 10 = *Much more irritating (pleasant) than the person drawn.* The wording of the question for the absolute evaluations of the target were similar to those used in Experiment 2. In the 5 or 2 vs. 1 conditions, the absolute rating of target was *If you had to give a score to the group of people as a whole, how would you score it?* Participants also provided an absolute evaluation of the referent

(1 person). For both absolute judgments, response scales similar to ones used in previous experiments were employed with appropriate adjustments of the anchors, as required by the condition. Again, the upper end of the scale always reflected more of the quality (pleasantness or irritation) under consideration. Finally, participants were debriefed and dismissed.

Results and discussion

Manipulation checks

As expected, participants' absolute evaluation of the target ($M = 8.6$, $SD = 1.1$) and referent ($M = 8.4$, $SD = 1.2$) were positive in the pleasant acquaintance condition, and significantly above the scale neutral point, p 's $< .001$. Similarly, absolute evaluations of the random target ($M = 7.8$, $SD = 1.7$) and referent ($M = 7.4$, $SD = 1.7$) in the irritating acquaintances condition were negative, $ps < .001$. Moreover, absolute ratings of single-item and multi-item entities did not significantly differ across conditions (all t s $< |1.7|$). Thus, in absolute terms, ratings of targets and referents did not differ.

Comparative judgments and relationships with absolute ratings

When participants compared five pleasant acquaintances to a single one, the B/TWA was not obtained ($M = -0.1$, $SD = 1.3$, $t(29) = -0.4$, ns). This result contrasts with the previous study where comparison of one pleasant acquaintance with five others led to the single target being rated as significantly more pleasant. By reversing the roles of the stimuli in Experiment 3, and thereby making the multi-item entity the target of comparison, the B/TWA was eliminated. Actually, there was slight trend to favor the referent, indicated by the negative sign. In the 2 vs. 1 condition, the mean comparative rating was marginally significantly negative ($M = -0.4$, $SD = 1.1$, $t(25) = -1.8$, $p = .09$), indicating that when two persons comprised the target, they were rated *less pleasant* (rather than more pleasant) than the 1 person referent—the opposite of the usual bias. These results suggest that making a multi-item entity the target undermines the usual bias. The trend for the multi-item target being rated as *less pleasant* than the single-item suggests more weight was given to the single-item, *even when it served as a referent.*

For the negative category, comparing five irritating people to one did not yield the typical comparative bias ($M = -0.5$, $SD = 2.5$, $t(29) = -1.2$, ns), nor did judging 2 vs. 1 ($M = -0.1$, $SD = 2.6$, $t(27) = -0.2$, ns). The sign of the means indicate that there was a trend for entities of five or two items to be rated as slightly *less irritating* than the single-item referent—another reversal of the usual effect. Means of the two referent size conditions did not significantly differ in magnitude in either the pleasant ($t < 1$) or irritating acquaintance conditions ($t < 1$).

In Experiments 1–3, the absence of an effect of number of items comprising the referent was interpreted as evidence against the group diffuseness account, but in this study the number of items *did* matter because there was no bias for a multi-item target. This result is counter to focalism, which predicts that any entity serving as the target, whether comprised of a single- or multiple-items, should be rated more extremely than any referent. The prediction from LOGE that the B/TWA bias would be obtained in both conditions also was disconfirmed. The UA-H prediction is that a multi-item target should not readily elicit a unique attribute so the target should not be rated more extremely than a referent. The results are consistent with that approach (and with the group diffuseness account). The trends for a (reverse) bias favoring the single-item referent also are noteworthy and consistent with the idea that a unique attribute should be easier to recruit for a single-item and may be used as the comparison standard underlying the B/TWA.

Zero-order correlations are presented in Table 4 for all four conditions. Recall that in Experiments 1 and 2, absolute ratings of the target consistently had stronger associations with the comparative ratings than ratings of the referent (see also Chambers, Windschitl, & Suls, 2003; Klar, 2002). This has been interpreted in past studies (e.g., Klar, 2002) as evidence that the target received more weight in the judgment. However, in this experiment where a multi-item entity served as the target, absolute ratings for the target and referent were both correlated with comparative ratings at roughly the same magnitude (in three of the four conditions). This suggests that there is a general tendency for multi-item entities to receive less weight regardless of serving as a target or a referent. Presumably, this is because of the difficulty of recruiting information about and evaluating multi-item entities (see Chambers & Windschitl, 2004).

In sum, the findings of Experiment 4 suggest that neither focalism nor LOGE can adequately account for multi-item vs. single-item comparisons. The absence of a bias was predicted from the perspective of both U-AH and group diffuseness, but the trend for a reverse bias (i.e., slightly favoring the single-member referent) is more consistent with U-AH. A general implication is that single-item entities (and their unique attributes) only receive more judgmental weight when they serve as targets rather than referents in the comparison.

Experiment 5

Multi-item entities appear to have a judgmental disadvantage in direct comparisons. However, in Experiments 1–4, a single-item always served either as the target, the referent, or both. This leaves open the possibility that focalism might confer an advantage when a multi-item entity is compared to another multi-item entity of the same size. However, if processing of a multi-item entity is cognitively demanding then it should not have a judgmental advantage, according to the group diffuseness account; also two multi-item entities should present difficulties in finding a unique attribute, according to U-AH. So neither the group diffuseness nor U-AH accounts predict that a multi-item entity should be rated more extremely than another multi-item entity. The focalism and LOGE accounts, on the other hand, predict that it should be rated more extremely. To test these different predictions, an experiment was conducted in which two entities comprised of the same number of items were compared.

Method

Participants and design

All participants within a given session were randomly assigned to either the pleasant or irritating acquaintances condition (com-

Table 4
Zero-order correlations between judgments as a function of acquaintance type and comparison set size in Experiment 3 (total $N = 114$).

	Pleasant acquaintances			Irritating acquaintances		
	Comp	Referent	Target	Comp	Referent	Target
<i>5 vs. 1 (N = 60)</i>						
Comp	–			–		
Referent	–.40*	–		–.43*	–	
Target	.32*	.07	–	.46*	–.08	–
<i>2 vs. 1 (N = 54)</i>						
Comp	–			–		
Referent	–.39*	–		–.55**	–	
Target	.60**	.34*	–	.50**	–.13	–

* $p < .10$.

* $p < .05$.

** $p < .01$.

paring stimuli on the left to right or right to left was counterbalanced, see below). Fifty-six undergraduate students participated.

Materials and procedure

Participants were asked to generate a list of either six pleasant or six irritating acquaintances. As before, each participant wrote the first names or initials of the six acquaintances on separate slips of paper and placed them in a large envelope. Next, each participant was asked to remove three slips and place them on the left side of the desk and to place the remaining three on the right side of the desk, after which they were presented with a questionnaire with three rating scales. When they finished making their ratings their materials were collected and they were debriefed and dismissed.

Measures

The three measures were similar to those used in the prior studies except both the target and referent referred to multi-item entities (each with three acquaintances). The direct comparison read: “Compared to the group of three on the left-side, how would you rate the group of three on the right?” followed by a 11-point scale with 0 = “much less pleasant (irritating) than the people on the right, 5 = “equally pleasant (irritating) to the people on the right,” and 10 = “much more pleasant (irritating) than the people on the right.” To counterbalance side, one-half of the subjects received the preceding instructions and rating scales; for the other half, the left and right labels were reversed so subjects compared the group on the right (target) to the group on the left (referent). Preliminary analyses indicated that the left–right position of the target had no effect so ratings were collapsed across that factor.

The absolute target measure was “If you had to give a score to the three people on the left (right), how would you score them as a group?” followed by a 0–10 point scale with 0 as “not at all pleasant (irritating)” and 10 as “extremely pleasant (irritating).” The absolute referent measure had the same form, “If you had to give a score to the three people on the right (), how would you score them as a group?” and same rating scale.

Results and discussion

Comparative judgments and relationships with absolute ratings

As in previous experiments, a t -test was performed to examine the presence of comparative bias. The mean rating in the pleasant acquaintances condition was 5.1 ($SD = 1.6$), which was not statistically different from the midpoint, one-sample $t < 1$. Likewise, the mean rating in the irritating acquaintances condition was 5.7 ($SD = 2.6$) and also not significantly different from the midpoint, one-sample $t(27) = 1.4$, ns . The means for these two conditions also did not differ from each other, $t(54) = 1.1$, ns . In short, when two multi-item entities were compared no significant judgmental advantage was found.

For the pleasant condition, absolute ratings of the three acquaintances located on the right side ($M = 8.6$, $SD = 1.2$) were significantly higher than those of the three acquaintances located on the left side ($M = 7.8$, $SD = 1.5$), $t = 2.4$, $p < .03$. This is likely a chance finding because members of the two entities were randomly selected and ratings were collapsed across left/right positions. Absolute ratings in the irritating condition did not significantly differ (left: $M = 5.9$, $SD = 2.5$; right: $M = 6.3$, $SD = 2.1$), $t < 1$.

Table 5 shows the correlations of absolute judgments of the two multi-item entities with the direct comparison ratings. In contrast to previous experiments where absolute assessment of a single-item target appeared to play a larger role in the direct comparison judgment, the contributions of the absolute ratings for multi-item entities (target and referent) were more comparable in magnitude. This is consistent with not finding the B/TWA effect. The evidence

Table 5

Z-order correlations between comparative ratings and absolute ratings (Study 4).

	Pleasant		Irritating	
Comp	–		–	
Target	.41*	–	.66*	–
Referent	–.42*	.12	–.38*	–

* Correlation significant, $p < .05$.

also indicates, as in Experiment 4, that a multi-item entity receives no judgmental benefit from serving as the comparison target. Also, as the U-AH and diffuse-group explanations predicted, no B/TWA emerged. Although focalism appears to operate for single-member targets, multi-item entities do not obtain increased judgmental weight when they serve as comparison targets.

Experiment 6

The final question we examined is whether a multi-item entity must always suffer a judgmental disadvantage in comparison. Our groups consisted of exemplars from ad hoc categories, such as healthy foods and irritating people. Researchers (McConnell, Sherman, & Hamilton, 1994) have shown that it is difficult to consider a set of exemplars as a cohesive entity for ad hoc groups, such as these kinds of categories. A lack of cohesiveness should make it difficult to both form an overall impression and retrieve information for the group's superiority (or inferiority in the case of a negative category). Thus, low cohesiveness militates against focalism and reinforces group diffuseness. Further, a lack of cohesiveness among items should make finding a unique attribute difficult. A manipulation that increases the perceived cohesiveness of a multi-item entity should both: (a) reduce the perceived diffuseness of that entity and (b) highlight feature(s) that are common for all items within that entity and distinct from other entities with which is compared (i.e., its unique attribute). Thus, under conditions of increased cohesiveness, B/TWA may be obtained even for multi-item entity serving as a target.

The final experiment was designed to test this idea by facilitating the perceived cohesiveness/unique attribute of the multi-item target. McConnell et al. (1994) found that special instructions to form a "cohesive impression" of exemplars increased the perception of group cohesiveness and thereby facilitated recall and impression formation. Similarly, we hypothesized that with greater perception of cohesiveness among members, judges should find it easier to form an impression of a multi-item entity and facilitate the recruitment of a unique attribute to serve as a comparison standard. Under such conditions, a multi-item entity serving as the target in the comparison should obtain the same judgmental advantage typically found for a target consisting of a single-item.

Experiment 6 involved two conditions, both of which were designed to increase perceived cohesiveness/unique attributes among items in the multi-item target. Both manipulations should increase the perceived cohesiveness and consequently the multi-item target should be rated more extremely than a single-item referent.

Method

Participants

Forty-two undergraduate students (in groups of one to eight) participated in this experiment.

Materials and procedures

Participants were randomly assigned to one of two conditions, both of which involved comparing a 5-item target with a single-item referent. All participants were initially instructed to think of

their six favorite films, write the name of each film on a slip of paper, then place all six slips in an envelope and shuffle them. At the next step, they were asked to remove five slips at random and place them on one side of their desk, then remove the remaining sixth slip and place it on the other side of the desk.

At this point, the instructions diverged. Following McConnell et al.'s (1994) procedure, in the "Impression" condition, participants were instructed to take a few minutes to "form a coherent impression about what this particular group of five films was like." Participants were then asked "form a coherent impression of the single film" on the other side of the desk. Participants were asked to form impressions about the target and referent so neither was given an attentional advantage.

Following the "impression" instructions, participants completed the dependent variables. Specifically, they were asked: "Compared to the film on the one side, how would you rate the group of films on the other side of the desk." Ratings were made on an 11-point scale with "much less pleasant than the film" (0), "equally pleasant (5) and "much more pleasant," (10) as verbal labels. Participants also were asked to provide absolute ratings of the group of five films and then the randomly selected film. Specifically, participants were instructed "...to give an average score to the group of five films," on an 11-point scale with "not at all pleasant" (0) to "extremely pleasant" (10)—this item was included as an absolute evaluation of the target. They also made an absolute evaluation of the referent (the single film) on the same scale. In the other condition ("absolute judgment first"), a different approach was used to facilitate group cohesiveness. After generating the six films and the random draw with one film on side of the desk and the five films on the other side, participants were asked to make an absolute evaluation of the five films on an 11-point scale with "not at all pleasant" (0) to "extremely pleasant" (10). Next, participants rated the referent (the single film) using the same scale. Finally, the direct comparison rating was presented with the same dependent variable used in the "Impression" condition. Thus, in this condition, the absolute rating of the group was made prior to the direct comparison to prompt the participant to form a collective impression of the multi-item target.

Results and discussion

For ease of interpretation, the scale ratings for the direct comparison were transformed from 0 to 10 to -5 to $+5$ (where 0 = same). The mean pleasantness score was 0.6 ($SD = 1.5$) in the "Impression condition," which was significantly different from 0, one-sample statistic $t(21) = 2.0$, $p = .05$. For participants who were instructed to form a cohesive impression of the five films, the multi-item target was rated more pleasant than the single-item referent. In the previous experiments, a multiple-item target received no judgmental advantage (Experiments 4 and 5), but after participants were prompted to think of the items as a cohesive entity, it was treated like a single-item.

The mean pleasantness rating for the "absolute judgment first condition" was 1.4 ($SD = 1.9$), $t(20) = 3.4$, $p < .003$. Those participants who rated the multi-item target on an absolute basis, prior to making a direct comparison, rated it more extremely. The results of both conditions support the idea that a set of ad hoc exemplars serving as a target will receive more weight than a single-item referent if judges are given preliminary instructions to think of the exemplars as a cohesive entity.

Supplementary analyses

For the "impression" condition, correlations between the absolute ratings and comparison ratings were computed. Although neither absolute target ($r = .24$, ns) nor absolute referent ratings ($r = -.28$, $p = .11$) were significantly correlated with comparison

ratings, trends were apparent and consistent with prior findings. The mean absolute ratings for the multi-item target and single-item referent were both 8.6 (SD 's = 1.5 and 1.8) and not significantly different, $t < 1$. As in the prior studies, the target and referent were not rated differently on an absolute basis but the target was rated more extremely than the referent in the direct comparison.

For the “absolute group rating first” condition, absolute target ratings were positively related to comparative ratings ($r = .33$, $p < .07$), and absolute referent ratings were negatively related to them ($r = -.20$, $p = .11$), consistent with there being a positive impact of the target and a negative impact of the referent on the comparison rating; there was only a trend for the multi-item target to have more influence. The mean absolute ratings for the multi-item target and single-item referent were 8.3 ($SD = 1.8$) and 7.3 ($SD = 2.2$), respectively and significantly different; paired $t(20) = 2.6$, $p < .01$. This is the only instance in the experimental series where absolute ratings were different. This divergence does not detract, however, from the main finding that facilitating a cohesive impression of an ad hoc set of exemplars reversed the usual judgmental disadvantage for a multi-item entity.

Additional analyses

For purposes of evaluating the four theoretical accounts, we sought to determine the strength of the B/TWA under each of the various conditions in our experiments (single-item vs. multi-item, target vs. referent). Average effect sizes were computed and are displayed in Table 6. There are three major findings of note. First, effects were stronger when single-item targets were compared with multi-item referents (1 vs. 2, 5, or 6 conditions: Mean $d = .65$) than with single-item referents (1 vs. 1 conditions: Mean $d = .48$). Second, small effects were found when multi-item targets were compared multi-item referents (3 vs. 3 conditions: Mean $d = .17$), and a small reversal of the effect was found when they were compared with single-item referents (2 vs. 1 and 5 vs. 1 conditions: Mean d 's = $-.20$ and $-.15$, respectively). Third, increasing the cohesiveness of a multi-item target (mean d 's = $.60$) produced effects similar to those when a single-item target was compared to a multi-item referent (mean $d = .65$).

General discussion

Predictions derived from four explanations for the B/TWA were tested. To summarize the evidence, the B/TWA tends to be exhibited whether a single-item target is compared to another single-item or multi-item referent. These results are most damaging to the diffuse-group explanation, but predictions from focalism also did not fare well for comparisons between multi-item targets and single-item referents or for comparisons between multi-item

targets and referents. Focalism predicts that a target comprised of a single- or multiple-items should receive more judgmental weight, but multi-item targets failed to show the B/TWA. Because LOGE assumes the same standard (i.e., a composite of the local and global standards) will be used independent of type of target and referent, B/TWA should be exhibited in all scenarios, which was not the case.

U-AH predicted the results best because multi-item entities consistently had a disadvantage as a target (or referent). This is explained by the fact that a group should not as readily elicit a unique attribute to define the salient comparison standard. The U-AH perspective also explains the inconsistency for 1 vs. 1 comparisons. In such a scenario, it is possible for unique attributes to be recruited for both the target and referent so they could compete. Also consistent with U-AH were the trends for a small reverse bias when a multi-item target was compared to a single-item referent. The U-AH suggests that more weight should be given to the unique attribute of the target than the referent, in which case a single-item target should have a slight judgmental advantage over a single-item referent; a single-item referent should have a small judgmental advantage over a multi-item target (producing a slight reversal of the typical bias)—both effects we demonstrated in the current studies (see Experiments 1–4).

The final study demonstrated that the B/TWA can be found for a multi-item entity if a cohesive impression can be formed of the items comprising that entity. The results also provide supportive evidence that diffuseness and difficulties associated with recruiting a unique attribute are responsible for the usual judgmental disadvantage associated with multi-item entities. This means that multi-item entities, such as stereotyped groups serving as comparison targets, should exhibit the B/TWA to the degree the group is perceived to be highly cohesive (Campbell, 1958; Hamilton & Sherman, 1996; McConnell, Sherman, & Hamilton, 1997). Ad hoc groups figure in many judgment scenarios, such as evaluation of job candidates and consumer products, so oftentimes entities comprised of multiple-items will receive a judgmental disadvantage and not show the B/TWA.

Each of the four explanations ascribe different critical processes for the B/TWA to emerge. However, the extent to which they are mutually exclusive still remains unclear. In particular, although the predictions from U-AH performed the best, focalism also seems to operate. Recall that results seemed most consistent with the idea that the UA of a target obtains more weight than the UA of a referent. This suggests there is some preferential processing of information in the foreground, namely for the comparison target. Where focalism seems to be in error is that any target, regardless of size, demonstrates the bias.

The present results also shed light on past results showing that including the self into a multi-person referent group does not block a B/TWA favoring another person serving as a single-item target

Table 6
Average effect size across conditions in Experiments 1–6.

	1 vs. 1	1 vs. 2	1 vs. 6 (5)	2 vs. 1	5 vs. 1	5 vs. 1 (cohesiveness)	3 vs. 3
Study 1-pleasant foods	0.85	0.92	0.96				
Study 2-pleasant acquaintances	0.29	0.56	0.65				
Study 2-irritating acquaintances	0.47	0.45	0.47				
Study 3-favorite films	0.31		0.53				
Study 4-pleasant acquaintances				-0.35	-0.08		
Study 4-irritating acquaintances				-0.04	-0.21		
Study 5-pleasant acquaintances							0.07
Study 5-irritating acquaintances							0.27
Study 6-favorite films-abs judg first						0.75	
Study 6-favorite films-impression						0.44	
Average effect size (d)	0.48	0.64	0.65	-0.20	-0.15	0.60	0.17

Note: Abs judg first = rated the target and referents on absolute rating scales prior to making direct comparison.

(Krizan & Suls, 2008). For example, people judged a friend to be “better than average,” even when comparing a friend with a referent group that *included the self*. These results were interpreted as support for focalism or group diffuseness and were clearly counter to egocentrism (i.e., self-relevant information is more influential than other-relevant information in comparison). But looked at in the context of the current findings, the simplest explanation is that the single target facilitated the recruitment of a unique attribute as a standard of comparison. Of course, it is noteworthy that even having the self in the referent, which could have served self-enhancement, did not overcome the tendency to favor the target.

The present results also reinforce earlier findings that patterns for direct comparisons and absolute ratings are strikingly different. Across experiments, there was a general tendency to judge the single-item more extremely than the referent in direct comparison, whereas absolute ratings of the target and referent did not differ. This is interesting because the difference score between absolute ratings of target and those of the referent have been used in some research as an index of comparative bias (e.g., Alicke, 1985; Brown, 1986). However, this indirect bias is weaker and more inconsistent than the direct comparison bias (Chambers & Suls, 2007; Chambers & Windschitl, 2004; Helweg-Larsen & Shepperd, 2001; Klar & Giladi, 1999).

The results also have relevance for Moore's (2007) view that B/WTA effects may be caused by differential regressiveness of target relative to referent judgments. He notes that absolute ratings of the target (in his examples, the self) will be higher than absolute ratings for the referent (in his examples, another person) on easy tasks, but lower on difficult tasks. This means that estimates of the target and referent are differentially regressive. “The natural consequence of the fact that people have more information about themselves than others is their self-assessments will be more extreme than will their assessments of others . . .” (Moore, 2007, p. 46).

Two aspects of our results are inconsistent with the differential regressiveness account. First, absolute ratings of the target and the referent did not significantly differ. Second, Moore posits that the variability associated with group judgments should be smaller than for individual targets. Across experiments, however, the variance in absolute judgments of single-item entities was not consistently larger than the variance in absolute judgments of multi-item entities. In fact, the variance in both judgments was often of approximately the same magnitude (Studies 1 and 3). Finally, Moore assumes participants have more information about the target than the referent. For self-other judgments this is the case, but the present paradigm employed ad hoc categories (and Giladi & Klar's, 2002 NSS/IBs) in which people should have equivalent information about the target and referent; in addition, the targets in our studies were chosen at random and therefore participants would have no more information about the target than the referent.

The findings also further clarify the relative roles of absolute ratings of target and referent for direct comparison. Most prior studies find that absolute ratings of the target are more strongly correlated with comparative ratings than are the absolute ratings of the referent (e.g., Klar, 2002; Kruger, 1999), which is taken as evidence that the target receives more weight in the comparative judgment. But in these experiments, the comparison target always consisted of a single-item. When similar situations were created in Experiments 1–2, comparable results were obtained—correlations between absolute referent and direct comparison ratings were smaller in magnitude and often statistically non-significant. However, when single-item referents were compared with multi-item targets (Experiment 4), the typical weighting advantage favoring the target disappeared. This pattern suggests more weight was given to a single entity (than a multi-item entity), even when it served as the referent and could not benefit from focalism. The

correlations in Experiment 5 suggest that when two multi-item entities are compared, the absolute ratings for both entities play approximately equivalent roles—leading to no bias for the target or the referent.

These findings have practical implications. As the results apply to members of both positive- and negative categories, a single-item in a positive category serving as a target should be liked more than other items in that category, but a single-item in a negative category should be liked less. When cohesiveness among items is low, a multi-item entity of positive exemplars should not be liked more and a multi-item entity of negative exemplars should not be liked less. This means that objects of judgment can be strategically placed in target or referent roles to increase or decrease the difficulty of forming an impression or the ability to recruit a unique attribute, thereby resulting in a desired judgmental outcome.

A more general implication of the findings concerns the “online” nature of comparative judgments. Just as attitudes appear to be context-sensitive judgments (e.g., Schwarz, 1999; Wilson, Lindsey, & Schooler, 2000) so too are comparisons (Tversky, 1977). Previous evidence indicates that whether an entity is the comparison target or referent makes a difference; the present results indicate that whether the entity is comprised of a single-item or multiple-items and serves as target or referent also affects the online comparison process and subsequent judgment. Strategic manipulation of the number of entities and which serve as comparison targets or referents may have significant implications for decision-making about the self, consumer products and political decisions. For example, whether a political candidate stands with fellow party members or as a singular candidate and is the target or referent of comparison should make a substantive difference in how the candidate is evaluated.

Acknowledgments

The research was partly supported by National Science Foundation Grant BCS 99-10592. The authors want to acknowledge the contributions of the members of the Comparison and Health in Psychology (CHIP) research group and the helpful suggestions of Paul Windschitl.

Reference

- 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.
- Alicke, M., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J., & Vredenburg, D. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology*, *68*, 213–226.
- Brown, J. D. (1986). Evaluations of self and others: Self-enhancement biases in social judgments. *Social Cognition*, *4*, 353–376.
- Burson, K., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology*, *90*, 60–77.
- Campbell, D. T. (1958). Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science*, *3*, 14–25.
- Chambers, J. R. (in press). Why the parts are better (or worse) than the whole: The unique-attributes hypothesis. *Psychological Science*.
- Chambers, J. R., & Suls, J. (2007). The role of egocentrism and focalism in the emotion intensity bias. *Journal of Experimental Social Psychology*, *43*, 618–625.
- 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.
- Chambers, J. R., Windschitl, P. D., & Suls, J. (2003). Egocentrism, event frequency, and comparative optimism: When what happen frequently is “more likely to happen to me”. *Personality and Social Psychology Bulletin*, *29*, 1343–1356.
- Codol, J. P. (1975). On the so-called superiority conformity of the self behavior: Twenty experimental investigations. *European Journal of Social Psychology*, *5*, 457–501.
- Eiser, J. R., Pahl, S., & Prins, Y. R. A. (2001). Optimism, pessimism, and the direction of self-other comparisons. *Journal of Experimental Social Psychology*, *37*, 77–84.
- Giladi, E. E., & Klar, Y. (2002). When standards are wide of the mark: Nonselective superiority and inferiority biases in comparative judgments of objects and concepts. *Journal of Experimental Psychology: General*, *131*, 538–551.

- Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review*, 103, 336–355.
- Helweg-Larsen, M., & Shepperd, J. (2001). Do moderators of the optimism bias affect personal or target risk estimates? *Personality and Social Psychology Review*, 5, 74–95.
- Hoorens, V. (1995). Self-favoring biases, self-presentation, and the self-other asymmetry in social comparison. *Journal of Personality*, 63, 793–817.
- Johnson, D. D. P. (2004). *Overconfidence and war: The havoc and glory of positive illusions*. Cambridge, MA: Harvard University Press.
- Klar, Y. (2002). Way beyond compare: Nonselective superiority and inferiority biases in judging randomly assigned group members relative to their peers. *Journal of Experimental Social Psychology*, 38, 331–351.
- Klar, Y., & Giladi, E. E. (1997). No one in my group can be below the group's average: A robust positivity bias in favor of anonymous peers. *Journal of Personality and Social Psychology*, 73, 885–901.
- Klar, Y., & Giladi, E. E. (1999). Are most people happier than their peers, or are people just happy. *Personality and Social Psychology Bulletin*, 25, 585–594.
- Klar, Y., Medding, A., & Sarel, D. (1996). Nonunique vulnerability: Singular versus distributional probabilities and unrealistic optimism in comparative risk judgments. *Organizational Behavior and Human Decision Processes*, 67, 229–245.
- Klein, W. M., & Weinstein, N. D. (1997). Social comparison and unrealistic optimism about personal risk. In B. P. Buunk & F. X. Gibbons (Eds.), *Health, coping, and well-being: Perspectives from social comparison theory* (pp. 25–61). Hillsdale, NJ: Erlbaum.
- Krizan, Z., & Suls, J. (2008). Losing sight of oneself in the above-average effect: When egocentrism, focalism and group-diffuseness collide. *Journal of Experimental Social Psychology*.
- Kruger, J. (1999). Lake Wobegon be gone! The “below-average effect” and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology*, 77, 221–232.
- Kruger, J., & Burrus, J. (2004). Egocentrism and focalism in unrealistic optimism (and pessimism). *Journal of Experimental Social Psychology*, 40, 332–340.
- Kruger, J., Windschitl, P. D., Burrus, J., Fessel, F., & Chambers, J. R. (2008). The rational side of egocentrism in social comparisons. *Journal of Experimental Social Psychology*, 44, 220–232.
- Malmendier, U., & Tate, G. (2005). CEO overconfidence and corporate investment. *Journal of Finance*, 60, 6.
- McConnell, A., Sherman, S. J., & Hamilton, D. L. (1994). On-line and memory-based aspects of individual and group target judgments. *Journal of Personality and Social Psychology*, 67, 173–185.
- McConnell, A. R., Sherman, S. J., & Hamilton, D. L. (1997). Target entitativity: Implications for information processing about individual and group targets. *Journal of Personality and Social Psychology*, 72, 750–762.
- Moore, D. A. (2007). Not so above average after all: When people believe they are worse than average and its implications for theories of bias in social comparison. *Organizational Behavior and Human Decision Processes*, 102, 42–58.
- Moore, D. A., & Kim, T. G. (2003). Myopic social prediction and the solo comparison effect. *Journal of Personality & Social Psychology*, 85, 1121–1135.
- Perloff, L. S., & Fetzer, B. K. (1986). Self-other judgments and perceived vulnerability to victimization. *Journal of Personality and Social Psychology*, 50, 502–510.
- Regan, P. C., Snyder, M., & Kassin, S. M. (1995). Unrealistic optimism: self-enhancement or person positivity. *Personality and Social Psychology Bulletin*, 20, 1073–1082.
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, 54, 93–105.
- Skade, D. A., & Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science*, 9, 340–346.
- Suls, J., Lemos, K., & Stewart, H. L. (2002). Self-esteem, construal, and comparisons with the self, friends, and peers. *Journal of Personality and Social Psychology*, 82, 252–261.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, 103, 193–210.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327–352.
- Wilson, T., Lindsey, S., & Schooler, I. (2000). A model of dual attitudes. *Psychological Review*, 107, 101–126.
- Windschitl, P. D., Conybeare, D., & Krizan, Z. (2008). Direct-comparison judgments: When and why above- and below-average effects reverse. *Journal of Experimental Psychology: General*, 137, 182–200.
- Windschitl, P. D., Kruger, J., & Simms, E. N. (2003). The influence of egocentrism and focalism on people's optimism in competitions: When what affects us equally affects me more. *Journal of Personality and Social Psychology*, 85, 389–408.