Journal of Experimental Psychology: General
2012, Vol. 141, No. 1, 124–133


0096-3445/11/$12.00 DOI: 10.1037/a0024006

# A Scale Distortion Theory of Anchoring

Shane W. Frederick
Yale University

Daniel Mochon
University of California, San Diego

We propose that anchoring is often best interpreted as a scaling effect—that the anchor changes how the response scale is used, not how the focal stimulus is perceived. Of importance, we maintain that this holds true even for so-called objective scales (e.g., pounds, calories, meters, etc.). In support of this theory of scale distortion, we show that prior exposure to a numeric standard changes respondents' use of that specific response scale but does not generalize to conceptually affiliated judgments rendered on similar scales. Our findings highlight the necessity of distinguishing response language effects from representational effects in places where the need for that distinction has often been assumed away.

*Keywords:* anchoring, context effects, mental representation, response scales, scaling effects

It is well established that perceptions and judgments are affected by the context of preceding or concurrent stimuli (Bevan & Turner, 1964; Brown, 1953; Campbell, Lewis, & Hunt, 1958; Di Lollo, 1964; Helson, 1964; Hunt, 1941; Krantz & Campbell, 1961; Parducci, 1965; Sherif, Taub, & Hovland, 1958). For example, water at room temperature feels warmer if one's hand was previously submerged in cold water, and the teeth of a person with dark skin look whiter than the teeth of person with fair skin.

We propose here that analogous contrast effects occur between numbers on a response scale—that the perceived magnitude of a number is affected by other numeric values on that scale with which it is compared. For example, 900 pounds seems larger if compared with 20 pounds and seems smaller if compared with 5,000 pounds.

This proposal seems uncontroversial, but its implications are not. First, it raises the question of whether any scale (pounds, inches, hours, calories, dollars, etc.) can be thought of as objective. Second, it suggests that many judgmental context effects, such as anchoring (e.g., Chapman & Johnson, 2002; Jacowitz & Kahneman, 1995; Strack & Mussweiler, 1997; Tversky & Kahneman, 1974), may be best interpreted as scaling effects. For instance, consider a demonstration of anchoring, in which the weight of a giraffe in pounds is judged to be lower if respondents are first asked whether it is heavier than 20 pounds than if they are first asked whether it is heavier than 5,000 pounds.

Anchoring effects are typically explained in terms of selective accessibility of anchor-consistent information, whereby respondents test whether the anchor might be the correct answer and remain biased by the information that best supports that tentative conjecture (Chapman & Johnson, 1999; Mussweiler, 2003; Mussweiler & Strack, 1999, 2000b; Strack & Mussweiler, 1997). For example, if first asked whether a giraffe weighs more or less than 20 pounds, the information retrieved that is most consistent with a giraffe being small becomes selectively accessible relative to contradictory information, reducing its perceived size and lowering estimates of its weight (see Figure 1B).

In contrast to this mechanism, our theory of scale distortion suggests that anchoring effects may also occur because of a shift in the use of the response scale itself and not because of any deeper change in the underlying representation of the target being judged. In the context of the 20-pound anchor, candidate responses—such as 1,000 pounds—are perceived as larger, which distorts the mapping between internal representations of massiveness and the numbers used to communicate these representations.[1] Because the numbers on the scale now correspond to larger representations, smaller numbers are sufficient to communicate the same unchanged belief of the giraffe's size (see Figure 1C).

Our theory of scale distortion makes three predictions that we test in this article:

1. If respondents are asked to select a stimulus that corresponds to a provided scale value (e.g., to name an animal that weighs 1,000 pounds), their response is contrasted away from provided standards. For example, a respondent who has just estimated the weight of a compara-

This article was published Online First July 18, 2011.

Shane W. Frederick, School of Management, Yale University; Daniel Mochon, Rady School of Management, University of California, San Diego.

Both authors contributed equally to this work. Authorship order was determined by a coin flip (heads). We thank Shahzeen Attari, Dan Bartels, Brett Boshco, Daylian Cain, Zoë Chance, Ravi Dhar, Nick Epley, Edward Hall, Daniel Kahneman, Nathan Novemsky, Daniel Oppenheimer, Daniel Read, Carol Shin, and Joe Simmons for comments on drafts of this article.

Correspondence concerning this article should be addressed to Shane W. Frederick, Yale University, School of Management, P.O. Box 208200, New Haven, CT 06520, or Daniel Mochon, University of California, San Diego, Rady School of Management, 9500 Gilman Drive #0553, La Jolla, CA 92093. E-mail: shane.frederick@yale.edu or dmochon@ucsd.edu

[1] Though we propose numeric contrast effects as the mechanism underlying scale distortion, the mapping between stimuli and scale values may be distorted in other ways. Our main purpose in this article is not to endorse a specific form of scale distortion but to make the broader claim that even objective scales can be distorted by context and that many contextual effects, such as anchoring, may be best regarded as scaling effects—changes in the use of the scale in which judgments are reported, rather than changes in the perception of the stimuli being judged.

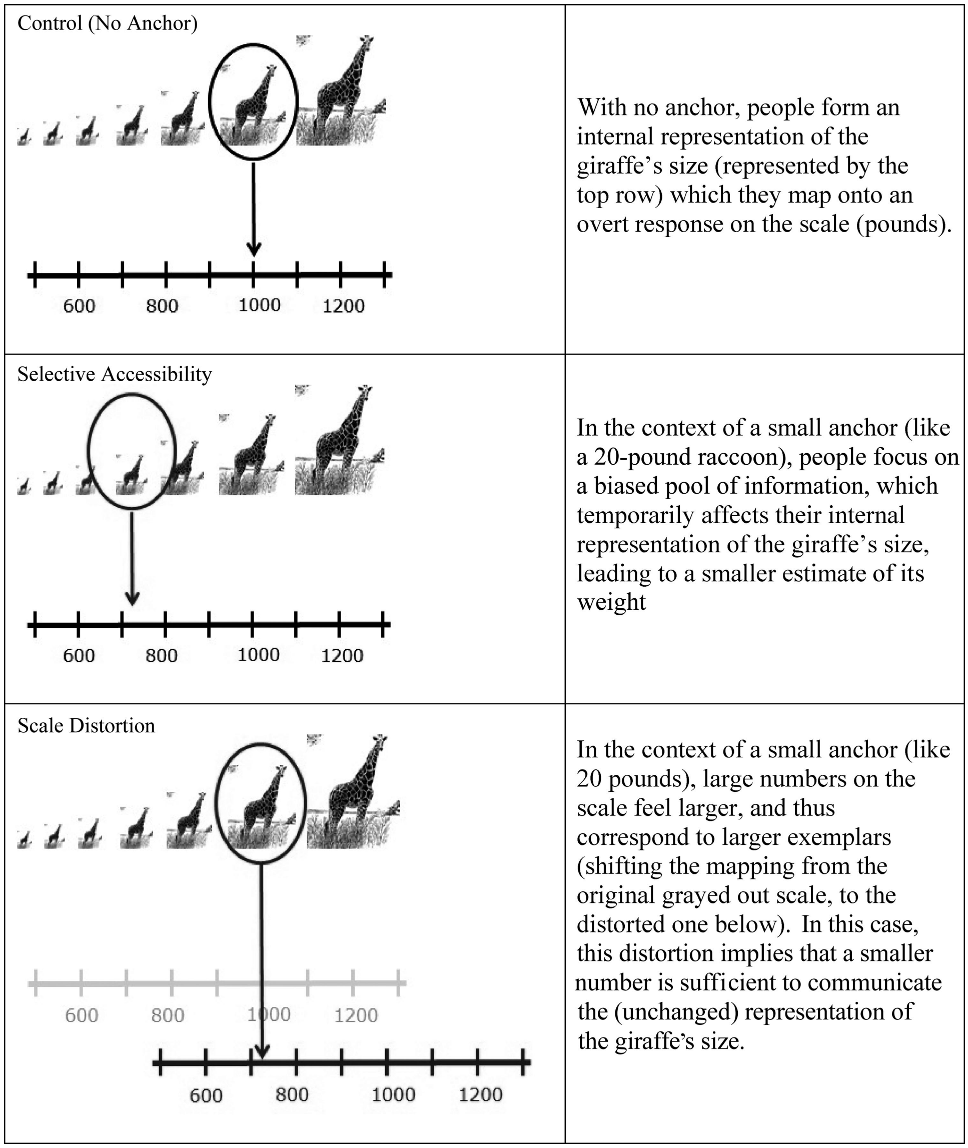


*Figure 1.* Two accounts of anchoring.

tively small animal (e.g., a 20-pound raccoon) should see 1,000 pounds as bigger by virtue of the comparison with the much smaller number they have just generated. Consequently, 1,000 pounds should correspond to a larger animal in the presence of this small anchor than without it. This prediction can be clearly seen in Figure 1C. Because of the contrast effect of the scale values, 1,000 pounds corresponds to a larger animal in the distorted bottom scale than in the original scale shown in gray above.

2. In more familiar tasks in which respondents select a scale value that corresponds to a provided stimulus (e.g., to estimate the weight of a giraffe in pounds), the contrast effects occurring at the scale level result in assimilation (or *anchoring*) on the overt response. For example, after having just estimated that a raccoon weighs 20 pounds, the three- and four-digit numbers that are candidate responses for a giraffe's weight seem larger by virtue of the comparison, and someone who might otherwise have selected the number 1,000 to represent their impression of a giraffe's size might now conclude that that a smaller value, like 700 pounds, is adequate (Figure 1C). Thus, our theory of scale distortion can explain anchoring effects without invoking any change in the underlying representation.

3. Because we propose that anchoring effects are often due to the distortion of a particular response scale rather than to a change in the representation of the judged stimulus, we predict that judgments about preceding stimuli will typically not affect the target judgment unless both judgments are

> made on the same response scale. As a consequence, if successive judgments are made on different scales, there will likely be no anchoring, even when the two judgments pertain to identical or conceptually affiliated judgmental dimensions. For example, though estimating the weight of a raccoon or a whale in pounds affects the subsequent judgment of the weight of a giraffe in pounds, we predict no such effect if the raccoon's weight is assessed on a 7-point heaviness scale, or if the whale's weight is assessed in tons. Furthermore, when anchoring effects do occur, they should be observed only on the specific scale that has been distorted and not on conceptually related scales that tap into the same representation.

The first prediction is supported by Studies 1, 2A, and 2B. The second and third predictions are supported by Studies 3 through 5. All six studies (and three others presented in the General Discussion section) support our notion that anchoring effects often reflect a change in respondents' use of the so-called objective response scale, rather than a change in their impression of the target stimulus. We conclude by discussing the tenable distinctions between subjective and objective scales, other evidence consistent with the view we propose, and the implications of this view for interpreting anchoring effects across various experimental paradigms.

## Study 1

### Method

One hundred and fifty-seven participants recruited from an online survey site examined a list of 15 animals ordered by weight, from very light (mouse) to very heavy (elephant) and selected the one whose average adult weight was closest to 1,000 pounds. Half of the participants estimated the weight of an average adult wolf before making this target judgment; the other half did not. Following the theory outlined earlier, we predicted that prior exposure to a smaller numeric standard would change respondents' interpretation of other scale values—specifically, that those in the wolf condition (who had just generated a much smaller number) would associate 1,000 pounds with a larger animal than those in the control condition (see Figure 1C).

### Results and Discussion

As predicted, the true weights of the creatures selected as examples of a 1,000-pound animal were significantly larger for the wolf group (2,170 pounds) than the control group (1,385 pounds); $t(156) = 2.26$, $p < .05$, $d = .36$.[2]

## Study 2A

To demonstrate that the result of the previous study extends to other stimuli and scales, we performed two conceptual replications in a different domain.

### Method

Two hundred and six participants received a brief paper-and-pencil survey listing 13 food items ordered from least caloric (hard-boiled egg) to most caloric (Burger King Whopper with cheese). Their task was to select the item closest to 400 calories. Half of the participants first estimated the number of calories in an average apple before making this judgment, whereas the other half did not.

### Results

The conceptual replication succeeded. Prior estimates of the number of calories in an apple appeared to increase participants' perception of what 400 calories means. Respondents in the apple condition picked items that were significantly more caloric, $M = 394$, $SE = 17.9$, than respondents in the control condition, $M = 330$, $SE = 17.5$; $t(204) = 2.5$, $p = .01$, $d = .35$.[3]

### Discussion

These results support our claim that prior judgments can affect respondents' interpretation of numeric labels even on objective scales, such as pounds and calories. Having just judged the weight of a wolf, participants selected a much larger representative for a 1,000-pound animal, and having just judged the number of calories in an apple, participants selected a more caloric example of a 400-calorie item.

These results support our theory of scale distortion, which proposes that even for objective scales, perceptions of the meaning of numeric scale values are affected by comparisons with other values on that scale. Further support is provided by the follow-up study reported next.

## Study 2B

### Method

Three hundred and five picnickers were shown a photo of a glass and asked to draw a line on it corresponding to the level that the glass would be filled by a 200-calorie serving of Hershey's chocolate syrup. Respondents were randomly assigned to either a control condition, in which only that judgment was requested, or a low-anchor condition, in which they first estimated the number of calories in a Hershey's kiss. We predicted that 200 calories would seem larger to those who first estimated the (much smaller) number of calories in a Hershey's kiss; therefore, we predicted that these respondents would indicate a glass with more syrup in it.

### Results

This prediction was confirmed. Respondents in the kiss condition drew a line 52.5 mm ($SE = 2.4$) above the bottom of the glass,

[2] The presented figures are geometric means. The data can also be analyzed by comparing the rank of the selected animals on the provided ordinal scale, which ranged from 1 (*mouse*) to 15 (*elephant*). Similar results were obtained. Respondents who first estimated the wolf's weight later chose larger animals, 11.6 versus 10.5; $t(156) = 2.01$, $p < .05$, $d = .32$.

[3] The pattern was replicated with the rank of the items selected. The mean rank of the food item selected by respondents in the apple condition ($M = 7.8$) was significantly higher than that of the one selected by respondents in the control condition ($M = 6.5$), $t(204) = 3.1$, $p < .01$, $d = .43$.

compared with 44.9 mm ($SE = 2.2$) for the control group, $t(340) = 2.29$, $p < .05$, $d = .26$.

## Study 3A

The prior studies used an atypical design in which respondents selected a stimulus that corresponded to a provided scale value. In those designs, the contrast effect is transmitted directly to the overt response (i.e., if the context increases the perceived magnitude of 1,000 pounds, respondents select a larger example of a 1,000-pound animal). The remaining studies use a more typical design in which respondents are provided with a stimulus and must produce a number. In these cases, prior numeric responses create a comparative standard against which to judge the magnitude of the chosen response. As a result, the contrast effect between scale values causes assimilation in the produced response (i.e., if 1,000 pounds seems larger in the context of a 23-pound raccoon, a smaller number may be sufficient to characterize the size of a giraffe). Of course, contrast requires comparison, and numbers that pertain to different scales are less likely to be directly compared. Thus, we further predict that anchoring effects will be limited to cases in which the contextual judgment and target judgment involve identical response scales; as we show, if the two scales differ, prior judgments do not affect the subsequent target judgment—even when the two scales pertain to the same judgmental dimension. This prediction follows from our claim that contextual effects result from a distortion of the perception of the numeric labels on the specific response scale, rather than a change in how the target stimulus is mentally represented.

### Method

Prior to the 4th of July fireworks display in Boston, 467 picnickers were randomly assigned to one of three conditions. The control group judged only the weight of a giraffe in pounds. The raccoon-pounds group first judged the weight of a raccoon in pounds ($M = 22$), and then judged the weight of a giraffe in pounds. The raccoon-heaviness group first judged the weight of a raccoon on a 7-point heaviness scale ($M = 3.56$), and then estimated the weight of a giraffe in pounds.

### Results

The mean estimates of the giraffe's weight were 1,254 pounds in the control group, 709 pounds for the raccoon-pounds group, and 1,265 pounds for the raccoon-heaviness group. Thus, prior judgments about a raccoon's weight induced a pronounced assimilation effect if rendered on the same scale as the target judgment, $t(320) = 4.42$, $p < .0001$, $d = .49$, but otherwise had no effect, $t(310) = 0.06$, $p > .9$, $d = .01$. These results support our theory of scale distortion. Because the two contextual conditions involved the same stimulus (a raccoon) and the same dimension (weight), our finding that the contextual effects are limited to the condition involving the identical response scale suggests that this is a scaling effect, rather than, say, a change in the relative accessibility of various pieces of information induced by a comparison between the giraffe and raccoon (Mussweiler, 2003). The similarity between the control condition and the raccoon heaviness condition also weighs against a numeric priming or conceptual priming interpretation of this effect (Critcher & Gilovich, 2008; Oppenheimer, LeBoeuf, & Brewer, 2008; Wilson, Houston, Etling, & Brekke, 1996; Wong & Kwong, 2000), as do the results presented next (see also Mochon & Frederick, 2011).

## Study 3B

Study 3A showed that even when the preceding judgment involved the same dimension as the target judgment (heaviness), there was no evidence of assimilation if the two judgments were rendered on different scales (a subjective 7-point heaviness scale vs. an unbounded objective number of pounds). That is, even though raccoons were judged as somewhat light (below the midpoint of the scale), the presence of this numeric value (which pertained to heaviness) did not affect subsequent judgments of the giraffe's weight. Here, we show that the preceding judgment does not affect the subsequent judgment if even the units differ—even though both judgments are rendered on objective scales that pertain to weight.

### Method

In the summer following Study 3A, in a new location with a new group of respondents, we randomly assigned 453 picnickers in Boston to one of three conditions. The target judgment in all three conditions was the weight of an average adult giraffe in pounds. The control group made only that judgment. The two other groups first judged the weight of an average adult blue whale, in either pounds ($M = 252{,}000$) or tons ($M = 56$).[4] Though these values are, respectively, well above and well below the numeric values appropriate for a giraffe judgment, we expected to find anchoring only among participants in the pounds group, who estimated weight on the same scale as the target judgment.

### Results

The mean estimates of the giraffe's weight were 1,053 pounds in the control group, 3,726 pounds for the whale-pounds group, and 1,155 pounds for the whale-tons group. Thus, prior judgments about a whale's weight affected subsequent judgments about a giraffe's weight when rendered in pounds, $t(303) = 1.99$, $p < .05$, $d = .23$, but not when rendered in tons, $t(301) = 0.62$, $p > .5$, $d = .06$. Once again, because the two contextual conditions involved the same stimulus (whale) and the same dimension (weight), our finding that judgmental effects are limited to the condition involving the identical scale suggests it is a scaling effect, rather than, say, the residue of attempts to test the hypothesis that a giraffe could weigh as much as a whale (Mussweiler, 2003). Moreover, the nonsignificant difference between the control condition and the whale-tons condition suggests that the anchoring effect for the whale-pounds group is not an example of numeric priming or basic anchoring (Wilson et al., 1996).

## Study 4

Studies 3A and 3B indicate that the anchor and target sharing the same response scale is a crucial component for the occurrence

[4] For the pounds judgments, the mean excludes two outliers whose estimates were (way) more than three standard deviations above the mean.

of anchoring. In this study and the next, we test this further, but manipulate the scales of judgment about the target rather than the anchor. Following Upshaw (1978), we propose that if the anchoring effect reflected a change in respondents' mental representation of the target stimulus, then it should extend to all conceptually affiliated judgments: For instance, if a prior judgment about a raccoon's weight makes giraffes seem smaller, then this should be reflected by all judgments related to their size, including height, weight relative to other objects, gestation period, the amount of vegetation consumed in a day, the number of people required to drag one, the number of lions it could feed, and so on. If no contextual effects are found on these other dimensions, it supports the interpretation of anchoring as a scaling effect, reflecting a contextual distortion of the use of a particular response scale, rather than a change in the perception or mental representation of the target stimulus.

## Method

We recruited 218 respondents from various locations in the Boston area to complete a brief paper survey. Half estimated the weight of a raccoon (in pounds), and half did not. All participants then estimated three features of a giraffe: its weight (in pounds), its height (in feet), and its weight relative to a grand piano, on a scale ranging from 1 (*piano is much heavier*) to 7 (*giraffe is much heavier*), with 4 indicating equivalence.

## Results and Discussion

As shown in Table 1, the giraffe was judged to weigh fewer pounds by those who had first judged the weight of a raccoon, 760 pounds versus 1,117 pounds; $t(213) = 2.4$, $p < .05$, $d = .33$. Notably, however, the raccoon had no appreciable effect on conceptually affiliated estimates rendered on other scales, including a giraffe's height, 17.3 feet vs. 17.5 feet; $t(213) = 0.18$ $p > .8$, $d = .02$, or its weight relative to a grand piano, 4.5 versus 4.5; $t(213) = 0.23$ $p > .8$, $d = .03$.

A follow-up study conducted as part of a larger survey with 148 picnickers in Boston yielded similar results. All participants estimated the number of adult lions that one adult giraffe could feed. One group made only that target judgment. A second did so after first estimating the weight of an adult giraffe, and a third did so after first estimating the weight of a wolf and a giraffe. Consistent with prior results, the preceding wolf judgment substantially depressed judgments of a giraffe's weight, 702 versus 1,095; $t(96) = 1.95$, $p = .05$, $d = .4$, but did not affect judgments of how many lions a giraffe could feed.

Table 1
*Mean Estimates for Study 4*

| Judgment 1 | Target judgment | | |
|---|---|---|---|
| Raccoon weight (lbs) | Weight of giraffe (lbs) | Height of giraffe (ft) | Weight of giraffe (vs. piano) |
| | 1,117 (130.3) | 17.5 (1.2) | 4.5 (0.15) |
| 22 (1.7) | 760 (72.3) | 17.3 (1.0) | 4.5 (0.11) |

*Note.* Standard errors are in parentheses.

Thus far, we have consistently found that judging the weight (in pounds) of much smaller animals substantially reduces judgments of a giraffe's weight (in pounds). However, we found no evidence that the mental representation of the giraffe is altered. By all metrics other than pounds, the giraffe appears to remain equally large in one's mind's eye—they remain as tall, as large relative to other objects, and as capable of nourishing hungry lions.

The lack of an effect on conceptually related metrics is especially striking because the overt pounds response, however it arose, could have been used as a basis for these other judgments (e.g., "Well, each lion could eat about 50 pounds of flesh, so if an 800 pound giraffe is half meat, it could feed about 8 of them"). A similar point is made by Sherman, Ahlm, Berman, and Lynn (1978): "[a change in] the scale rating of the target issue, *whatever its basis*, may serve as an input for cognitions about the target, and thus may be a *cause* of changes in content and meaning, rather than an *effect* of these changes" (p. 342). Thus, though the presence of contextual effects on judgments of conceptually affiliated attributes would not rule out scale effects, the failure to find such effects surely weighs against the idea that anchoring of the numeric response is accompanied by broader changes in the mental representation of the target stimulus. The following study shows this in another domain.

# Study 5

## Method

After random assignment to one of two conditions, 147 respondents recruited from an online survey site estimated either the number of calories in a strawberry (mean estimate = 28) or the number of calories in a large Domino's cheese pizza (mean estimate = 1,717). Both groups then made five judgments about a medium serving of McDonald's fries: (a) the number of calories it contains, (b) the number of grams of fat it contains, (c) the number of pounds that could be lost if the participant quit eating fries for a year, (d) the number of fries in a serving, and (e) the number of days a rat could survive on the specified serving.[5]

If the context provided by the preceding judgment (about a strawberry or cheese pizza) affects respondents' conception of McDonald's fries, we would expect these other judgments to also be affected. However, if the context effect is limited to respondents' use of the calories scale, judgment on these other attributes may be unaffected.

## Results and Discussion

The results support a scaling effect interpretation. As Table 2 shows, a medium serving of McDonald's fries was judged to have significantly fewer calories by the strawberry group than by the pizza group, 261 versus 379; $t(145) = 3.3$, $p < .001$, $d = .55$. However, respondents did not judge a serving of fries as having less fat, $t(145) = 1.3$, $p > .15$, $d = .22$, being a less important consideration for dieters, $t(145) = 0.3$, $p > .75$, $d = .05$, being

[5] Because of a programming error, 95 participants answered Question 4 but not Question 5, whereas 52 answered Question 5 but not Question 4. This error was independent of condition.

Table 2
*Mean Estimates for Study 5*

| Condition | Target judgment: Calories in medium fries | Grams of fat in medium fries | Pounds lost from not eating | No. of fries in medium fries | Days a rat could survive | Composite |
|---|---|---|---|---|---|---|
| Calories in strawberry: 28 (2.7) | 261 (16.3) | 40 (7.8) | 14 (1.5) | 47 (10.7) | 19 (6.0) | 0.01 (0.1) |
| Calories in pizza: 1,717 (311.4) | 379 (30.1) | 29 (3.9) | 13 (1.3) | 56 (6.9) | 27 (4.9) | −0.01 (0.1) |

*Note.* Standard errors are in parentheses.

smaller, $t(93) = 0.74$, $p > .45$, $d = .15$, or being less capable of nourishing a rat, $t(49) = 1.1$, $p > .25$, $d = .3$. Indeed, a composite $z$ score of these other caloric indices revealed that the low or high anchor had no effect on these other judgments, 0.01 versus −0.01; $t(145) = 0.1$, $p > .9$, $d = .01$.

## General Discussion

If prior exposure to an especially bright room causes respondents to describe a subsequent room as darker, it is important to know whether respondents just call it darker or whether they experience it as darker, as would be evidenced by them bumping into things they would normally avoid. On subjective scales, the distinction between response scale effects and real perceptual effects has been extensively discussed (e.g., Campbell et al., 1958; Helson, 1964; Parducci, 1965; Ross & Di Lollo, 1968, 1971; Stevens, 1958; Upshaw, 1978, 1984). The proper interpretation is sometimes clear. As Stevens (1958) quipped, one who reports seeing "a large mouse run up the trunk of a small elephant" (p. 633) is not confusing the relative size of the two animals but is using the labels *large* and *small* in a contextually sensitive way. However, other cases are more ambiguous. For a specific example, consumer researchers agree that lowering expectations about a product may increase customer satisfaction ratings but nevertheless disagree about the appropriate inference. Some conclude that customers who expect less have a better experience (Oliver, 1980; Parasuraman, Zeithaml, & Berry, 1985, 1988; Woodruff, Cadotte, & Jenkins, 1983), whereas others interpret this as a pure response-scale effect with minimal implications for actual behavior (see Lynch, Chakravarti, & Mitra, 1991).

Contextual effects observed on objective response scales (inches, pounds, dollars, etc.) have been less controversial because such scales are widely held to be immune to contextual shifts in meaning (e.g., Biernat, Manis, & Nelson, 1991; Campbell et al., 1958; Krantz & Campbell, 1961; Lynch et al., 1991; Mussweiler, 2003). Consider these excerpts:

> Objective judgments, however, are unlikely to be influenced by a salient reference point. Because judgments [on objective scales] are externally anchored by consensual standards, the underlying scale cannot be shifted. (Mussweiler & Strack, 2000a, p. 25)

> Objective response scales . . . do not allow for such shifts in meaning. This is because objective judgments are externally anchored. . . . For example, when judging targets' heights in feet and inches, we know that an inch is an inch. (Biernat, Manis, & Kobrynowicz, 1997, p. 256)

> The linguistic convention that allows words such as "important" or "satisfying" to be understood in a relative sense does not apply to [objective scales]. Consider the question: "What is the size of an eagle, in meters?". . . Of course, there is no reason to expect any [response scale] effect on answers to this question. *A context effect on a size judgment expressed in absolute units indicates* a visual illusion—*a change in the underlying perception, not in the language used to describe it* . . . [it indicates] that the evaluation itself, not only the expression of it, is altered by the comparison. (Kahneman, Ritov, & Schkade, 1999, p. 219, emphasis added)

In short, context effects on objective response scales are usually interpreted as reflecting changes in the underlying perception of the target stimulus, and their use is often advocated for this reason (e.g., Biernat et al., 1991; Krantz & Campbell, 1961; Mussweiler, 2003). For example, if told that a badger was rated as a 7 on a 10-point ferocity scale, we'd likely want to know with which other animals it was being compared. By contrast, if told that the mean estimate of a badger's weight was 27 pounds, we would not typically inquire about the context for that judgment. Correspondingly, if badgers are judged to weigh more pounds in one experimental condition than another, it is generally interpreted to mean that respondents' perception of badgers has (at least temporarily) been altered.

We challenge this presumption. Though all of our studies involve so-called objective scales (pounds, calories, etc.), we find no evidence that contextual effects on overt responses are accompanied by a broader change in respondents' conception of the judged stimulus. We propose instead that the observed anchoring effects are best interpreted as response language effects and that such effects play a much larger role than is customarily acknowledged. Our theory does not exclude the possibility that other processes could also occur, but it competes with other anchoring accounts in the sense that its explanatory sufficiency reduces the need to postulate anything else.

### Subjective and Objective Scales

Though we contend that so-called objective scales are susceptible to the same sort of contextual effects that affect subjective scales, distinctions do remain. Consider these two questions:

1. How heavy is an adult badger? ___ pounds

2. How heavy is an adult badger on a scale from 0 (*not at all*) to 10 (*extremely*)? ___

Two respondents giving the same answer to Question 1 may answer Question 2 differently, if different contexts are invoked for interpreting the meaning of *extremely heavy*. Thus, whereas the first response (on the objective scale) can be compared with a true value, the second (on the subjective scale) cannot, because the scale's endpoints and units can be interpreted as you wish.

Response scale effects on objective scales may also have different behavioral consequences than response scale effects on subjective scales. For example, if respondents rate a fancy bottle of wine as more attractive on a 10-point scale after comparing it with a pencil than after comparing it with an iPod, the context effect may represent only the judgmental relativity of labels. However, if respondents would actually pay less, on a dollars scale, after having just bid on a pencil than after having just bid on an iPod, evaluation order would affect how much money people actually spend and how much good wine they actually consume—indeed, Mochon and Frederick (2011) report such a result. Similarly, the finding that a medium serving of fries is labeled as *very caloric* in the context of a strawberry but merely *somewhat caloric* in the context of a pizza would certainly not imply a differential tendency to consume or forego those fries. However, a context effect on judgments of how many calories those fries contain might display such a correspondence—if, say, someone were literally counting calories and deciding whether those fries would put their daily total above their 2,000-calorie limit. As a final example of why responses on objective scales might matter, irrespective of whether they accompany changes in mental representations, suppose the objective length of a criminal's sentence is reduced by the assignment of a much smaller number in a preceding case. The victim's family would hardly take comfort from the news that this was just a response scale effect and not a true change in the judged severity of the penalized act.

Thus far, we have treated the distinction between subjective and objective scales as a dichotomy—objective scales (such as pounds) have a meaningful zero point, whereas subjective scales (such as heaviness) do not. However, some scales do not fit neatly in either category. Consider again an expression of willingness to pay, in dollars. The dollar scale is objective in several senses: It has a meaningful zero point, an unrestricted set of potential responses, and a consensual definition; everyone told that a ticket to the zoo costs $7 knows what 7 means. However, monetary responses are nevertheless quite subjective in other senses. If a person claims he would pay $70 to prevent the extinction of some rare bird, the response cannot be easily judged as right or wrong—not even by the person making it. Thus, it should not be surprising that responses on a dollar scale are, like responses on any scale, sensitive to contextual effects. For example, Ariely, Loewenstein, and Prelec (2003) showed that valuations for goods are sensitive to arbitrary anchors, and Kahneman, Schkade, and Sunstein (1998) discussed the spectacular arbitrariness of responses when respondents are asked to play jurors by translating their subjective impression of moral outrage onto a dollar scale of punitive damages.

## Implications for the Interpretation of Anchoring Effects

The widespread presumption that objective scales prevent or minimize scaling effects dovetails with a prevailing assumption that anchoring effects (which typically involve judgments on objective scales) reflect changes in respondents' conception of the judged object. According to the selective accessibility account of anchoring, the finding that judgments about a giraffe's weight (in pounds) are smaller following judgments about the weight of a raccoon is interpreted to mean that the giraffe is—at least momentarily—perceived as being smaller. By this account, respondents are thought to first test whether a giraffe could be as light as a raccoon, and though they quickly reject that low value as a candidate answer, they remain influenced by the most consistent facts conjured during that confirmatory search (e.g., that giraffes are skinny and have small heads), which have become comparatively accessible relative to more discrepant facts (e.g., that giraffes are very tall and capable of killing adult lions with one kick). For more nuanced articulation of the selective accessibility account, see Chapman and Johnson (1994, 1999); Mussweiler (2003); Mussweiler and Strack (1999, 2000b); and Strack and Mussweiler (1997).

Selective accessibility is a compelling account when respondents have a large pool of relevant knowledge from which to select. For example, in a study by Englich and Mussweiler (2001), respondents read about a potentially criminal act and received a four page summary containing various discrete pieces of aggravating and mitigating evidence. Prior to their judgment of the appropriate sentence length, respondents were exposed to a numeric anchor (another person's recommendation of the appropriate sentence in months). Here it seems quite plausible that the presented anchor might guide respondents' attention to the presented facts most consistent with a short or long sentence. However, within our paradigm and examples, the selective accessibility account seems less plausible. In particular, we doubt respondents' knowledge is sufficiently elaborated to provide much for selective accessibility to act upon. We suspect that respondents know that giraffes are fairly big and live in Africa but little else.

In the studies presented here, no comparative judgment was explicitly requested, which differentiates the paradigm we used from most other studies of anchoring. However, results from three other studies which use the standard paradigm also seem more supportive of scale distortion than selective accessibility. In one study conducted with students in a master of business administration program at the Massachusetts Institute of Technology, respondents first guessed whether the high temperature in Atlanta exceeded 80 degrees more than or fewer than 2/200 days a year, before estimating the average number of days per year it snows in Minneapolis. As might be anticipated, the snowfall judgment was significantly higher in the high anchor condition than in the low anchor condition, 51 days versus 36 days; $t(66) = 2.8$; $p < .01$, $d = .69$. However, if one adopts the selective accessibility view, it seems very hard to construct a story about how information consistent with Atlanta being hotter should cause respondents to think that Minneapolis is snowier.

In a similar study involving a 2 × 2 between-subjects design, we manipulated both the anchor value and the anchor stimulus of the comparative judgment. Specifically, participants in this study were asked whether an adult wolf/Ridley's sea turtle weighs more or less than 10/1,000 pounds, before rendering their best estimate of the weight of an adult wolf. The estimates showed a significant effect of anchor value, $F(1, 141) = 19.4$, $p < .001$, $\eta^2 = .12$, but no main effect of the anchor stimulus, $F(1, 141) = 0.21$, $p > .5$, $\eta^2 = .00$, and no interaction, $F(1, 141) = 0.02$, $p > .8$, $\eta^2 = .00$. Specifically, the wolf was judged to weigh significantly less in the low-anchor condition than the high-anchor condition, irrespective of whether the comparative judgment involved the wolf, 80 versus 195 pounds; $t(72) = 3.9$, $p < .01$, $d = .9$, or the sea turtle, 95 versus 203 pounds; $t(69) = 2.6$, $p < .05$, $d = .6$. Because the selective accessibility model attributes the anchoring effect to target-relevant knowledge rendered accessible during the comparative judgment (Mussweiler, 2003), shouldn't it matter whether the activated knowledge actually pertains to the target? To us, it seems highly odd to suppose that the most accessible facts about sea turtles should have a marked effect on judgments about wolves.

In a third study, 1,223 members of an online survey site were randomly assigned to one of 10 conditions. All participants were first told (truthfully) that the driest place in the United States is Death Valley, California, and that the wettest place is Mount Waialeale on the island of Kauai in Hawaii (hereafter referred to as DV or MW). Two calibration groups made a single unanchored judgment, simply giving their best estimate of the annual rainfall (in inches) in DV or MW. The remaining eight groups were composed using a 2 × 2 × 2 between-subjects design, which orthogonally manipulated locale of the comparative judgment (MW or DV), comparative value (2 inches or 460 inches), and the locale of the continuous target judgment (MW or DV).[6]

As shown in Table 3, we find again that target judgments are anchored on the values used in the comparative judgments, even when the comparative stimulus bears little relation to the target stimulus—indeed, even when those two stimuli are, by design, maximally different. Indeed, on the whole, the applicability of the comparative judgment appears not to matter at all. Anchoring effects are essentially identical, by two different indices: the difference between the median judgment of the high and low anchor groups divided by the difference between the high and low anchors (see Jacowitz & Kahneman, 1995) and the percentage of the responses in the low and high anchor groups that lie above the median judgment of the corresponding calibration group.

This study and the previous two collectively weigh against Strack and Mussweiler's (1997) claim that anchoring effects depend critically on the applicability of the comparative judgment to the absolute target judgment. Certainly, we find no evidence for their assertion that "anchors may exert directionally different effects depending on whether the object of the comparative task is identical with or distinctly different from the object of the absolute judgment" (Strack & Mussweiler, 1997, p. 441).

Our account suggests that the demonstration of anchoring effects—even on objective scales—need not betoken a deeper change in the mental representations of, or beliefs about, or attitudes toward the entities being judged. We are not the first to make this observation. Consider an unpublished study by Kahneman and Knetsch (1993). In a 2 × 2 between-subjects design, Toronto residents were first asked whether they would pay $25 (or $200) to protect lakes in the area from acid rain and were then asked to estimate the average amount that Toronto residents would be willing to pay (or the proportion willing to pay $100). Though the higher dollar anchor significantly increased judgments of mean willingness to pay, it had no effect on percentage judgments. As they noted, it would be odd to say that the anchor shifted respondents' beliefs about the distribution of willingness to pay if only one of two conceptually affiliated metrics shows an effect.[7] Brewer, Chapman, Schwartz, and Bergus (2007) implicate a form of scale distortion as a possible mechanism in a study in which HIV-positive gay men estimated the probability they would transmit their disease to a sexual partner if their condom broke during intercourse. Following the traditional anchoring paradigm (in which a comparative question precedes a continuous judgment), respondents first judged whether the probability exceeded 1% (or 90%), before rendering their best estimate of the exact probability. As might be expected, those in the higher anchor condition gave significantly higher estimates of the risk of transmission. However, despite this, the high anchor group did not recommend different actions if such an accident occurred (which ranged from doing nothing to commencing drug treatment). A further example is provided by Ariely et al. (2003) who showed that a monetary anchor value affected how much money participants would demand to endure an annoying sound, without affecting how annoying they considered that sound to be relative to other aversive experiences.

Such results are compatible with our view of anchoring as a scaling effect—a change in the way numeric labels are being used in a particular context unaccompanied by any change in the beliefs or mental representation of the entity being judged. The results seem less consistent with a selective accessibility account as it is typically adumbrated.

Our experimental results also readily distinguish our scale distortion theory from various anchoring and adjustment accounts, because only scale distortion predicts the results found across both types of experimental designs we used: those in which the participant selects the stimulus corresponding to a provided scale value (Studies 1, 2A, and 2B) and the more typical designs in which the participant chooses the scale value corresponding with a provided stimulus. Within this second paradigm, however, the two accounts can be difficult to distinguish, both empirically (because both predict assimilation toward a provided standard) and conceptually (because what is perceived as a reasonable stopping point in the adjustment along the scale is presumably affected by the perceived magnitude of the numbers over that region of the scale). Moreover, because both accounts share a focus on the response scale, they would be grouped together when being contrasted with theories, such as selective accessibility, that focus on changes in the mental representation of the stimulus being judged.

We should note that because our theory relies on contrast effects at the scale level, it likely has weak effects in instances in which the anchor and target are quite close (much as the Ebbinghaus illusion would be attenuated if the peripheral circles were nearly identical to the size of the central circle). Thus, we expect that scale distortion would have a negligible effect for most of the items within the self-generated anchor paradigm (Epley & Gilovich, 2001, 2004, 2005, 2006), which is restricted to judgments in which a strongly associated (and clearly relevant) referent is spontaneously brought to mind. However, even in this paradigm, scale distortion may occasionally help explain the insufficiency of adjustments. For example, if asked to estimate the freezing point of vodka, 32 °F would immediately come to mind for many people because it is the only freezing point they know. On reflection, this value would continue to be judged highly relevant (after all, vodka is mostly water), though many respondents would also recall that alcoholic beverages typically do not freeze solid in their freezers and would, therefore, restrict their consideration to values below 32 °F. To the extent that a considered response (say 5 °F) appears

[6] Though nothing was said about the source of these numbers, these are the actual values of mean rainfall in these two extreme locations.

[7] To make an analogy, a claim that women in Denmark are as tall as men could not be taken seriously if the claimant conceded that more Danish men than Danish women exceed any specified height.

Table 3
*Mean Estimates for Rainfall Study*

| Condition | *N* | Comparative judgment (% guessing more) | Target of absolute judgment | Median estimate (target) | 10% trim mean estimate | JK index | % distribution > median of calibration group |
|---|---|---|---|---|---|---|---|
| 1 | 125 | DV, 2 (30%) | DV | 1 | 2 | 0.02 | 26 |
| 2 | 114 | DV, 460 (7%) | DV | 10 | 69 | | 72 |
| 3 | 125 | MW, 2 (98%) | DV | 2 | 6 | 0.02 | 30 |
| 4 | 122 | MW, 460 (46%) | DV | 12 | 91 | | 72 |
| 5 | 125 | — | DV | 3 | 11 | — | — |
| 6 | 120 | MW, 2 (98%) | MW | 36 | 58 | 0.90 | 20 |
| 7 | 128 | MW, 460 (48%) | MW | 450 | 438 | | 93 |
| 8 | 128 | DV, 2 (35%) | MW | 30 | 61 | 0.94 | 28 |
| 9 | 114 | DV, 460 (9%) | MW | 460 | 396 | | 85 |
| 10 | 122 | — | MW | 98 | 140 | — | — |

*Note.* JK = Jacowitz and Kahneman (1995); DV = Death Valley, California; MW = Mount Waialeale on the island of Kauai in Hawaii.

colder by virtue of comparison with the chronically accessible numeric referent (32 °F), respondents would choose numbers closer to that referent (i.e., anchoring effects would be stronger).[8]

## Conclusion

We propose that anchoring is often best interpreted as a scaling effect and that contextual effects on objective scales cannot be assumed to reflect changes in respondents' conception of the target stimulus. Our findings resurrect the challenge of distinguishing representational effects from response language effects across a broad array of judgments where the need for that distinction has often been assumed away. Brass weights and other physical standards can be safeguarded at the international bureau of weights and measures, but the psychological meaning of 1,000 pounds or 1,000 calories is less easily protected.

[8] One of the limitations of the self-generated anchoring paradigm is that there is no obvious way to create a control group by which to define the amount of anchoring that is occurring. In the vodka example, one would have to identify a group of people who do not know the freezing point of water, yet who are nevertheless cognitively equivalent to those who do. There may be no such group.

## References


Ariely, D., Loewenstein, G., & Prelec, D. (2003). "Coherent arbitrariness": Stable demand curves without stable preferences. *Quarterly Journal of Economics, 118,* 73–105. doi:10.1162/00335530360535153

Bevan, W., & Turner, E. D. (1964). Assimilation and contrast in the estimation of number. *Journal of Experimental Psychology, 67,* 458–462. doi:10.1037/h0041141

Biernat, M., Manis, M., & Kobrynowicz, D. (1997). Simultaneous assimilation and contrast effects in judgments of self and others. *Journal of Personality and Social Psychology, 73,* 254–269. doi:10.1037/0022-3514.73.2.254

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

Brewer, N. T., Chapman, G. B., Schwartz, J. A., & Bergus, G. R. (2007). The influence of irrelevant anchors on the judgments and choices of doctors and patients. *Medical Decision Making, 27,* 203–211. doi:10.1177/0272989X06298595

Brown, D. R. (1953). Stimulus-similarity and the anchoring of subjective scales. *American Journal of Psychology, 66,* 199–214. doi:10.2307/1418726

Campbell, D. T., Lewis, N. A., & Hunt, W. A. (1958). Context effects with judgmental language that is absolute, extensive, and extra-experimentally anchored. *Journal of Experimental Psychology, 55,* 220–228. doi:10.1037/h0045310

Chapman, G. B., & Johnson, E. J. (1994). The limits of anchoring. *Journal of Behavioral Decision Making, 7,* 223–242. doi:10.1002/bdm.3960070402

Chapman, G. B., & Johnson, E. J. (1999). Anchoring, activation, and the construction of values. *Organizational Behavior and Human Decision Processes, 79,* 115–153. doi:10.1006/obhd.1999.2841

Chapman, G. B., & Johnson, E. J. (2002). Incorporating the irrelevant: Anchors in judgments of belief and value. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 120–138): New York, NY: Cambridge University Press.

Critcher, C. R., & Gilovich, T. (2008). Incidental environmental anchors. *Journal of Behavioral Decision Making, 21,* 241–251. doi:10.1002/bdm.586

Di Lollo, V. (1964). Contrast effects in the judgment of lifted weights. *Journal of Experimental Psychology, 68,* 383–387. doi:10.1037/h0042094

Englich, B., & Mussweiler, T. (2001). Sentencing under uncertainty: Anchoring effects in the courtroom. *Journal of Applied Social Psychology, 31,* 1535–1551. doi:10.1111/j.1559-1816.2001.tb02687.x

Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science, 12,* 391–396. doi:10.1111/1467-9280.00372

Epley, N., & Gilovich, T. (2004). Are adjustments insufficient? *Personality and Social Psychology Bulletin, 30,* 447–460. doi:10.1177/0146167203261889

Epley, N., & Gilovich, T. (2005). When effortful thinking influences judgmental anchoring: Differential effects of forewarning and incentives on self-generated and externally provided anchors. *Journal of Behavioral Decision Making, 18,* 199–212. doi:10.1002/bdm.495

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic. Why the adjustments are insufficient. *Psychological Science, 17,* 311–318. doi:10.1111/j.1467-9280.2006.01704.x

Helson, H. (1964). *Adaptation-level theory: An experimental and systematic approach to behavior.* New York, NY: Harper & Row.

Hunt, W. A. (1941). Anchoring effects in judgment. *American Journal of Psychology, 54,* 395–403. doi:10.2307/1417684

Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin, 21,* 1161–1166. doi:10.1177/01461672952111004

Kahneman, D., & Knetsch, J. L. (1993). *Anchoring or shallow inference: The effect of format.* Unpublished manuscript.

Kahneman, D., Ritov, I., & Schkade, D. A. (1999). Economic preferences or attitude expressions? An analysis of dollar responses to public issues. *Journal of Risk and Uncertainty, 19,* 203–235. doi:10.1023/A:1007835629236

Kahneman, D., Schkade, D. A., & Sunstein, C. R. (1998). Shared outrage and erratic awards: The psychology of punitive damages. *Journal of Risk and Uncertainty, 16,* 49–86. doi:10.1023/A:1007710408413

Krantz, D. L., & Campbell, D. T. (1961). Separating perceptual and linguistic effects of context shifts upon absolute judgments. *Journal of Experimental Psychology, 62,* 35–42. doi:10.1037/h0040386

Lynch, J. G., Chakravarti, D., & Mitra, A. (1991). Contrast effects in consumer judgments: Changes in mental representation or in the anchoring of rating scales. *Journal of Consumer Research, 18,* 284–297. doi:10.1086/209260

Mochon, D., & Frederick, S. (2011). *Anchoring in sequential judgments* (Working paper). New Haven, CT: Yale University.

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

Mussweiler, T., & Strack, F. (1999). Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model. *Journal of Experimental Social Psychology, 35,* 136–164. doi:10.1006/jesp.1998.1364

Mussweiler, T., & Strack, F. (2000a). The "relative self": Informational and judgmental consequences of comparative self-evaluation. *Journal of Personality and Social Psychology, 79,* 23–38. doi:10.1037/0022-3514.79.1.23

Mussweiler, T., & Strack, F. (2000b). The use of category and exemplar knowledge in the solution of anchoring tasks. *Journal of Personality and Social Psychology, 78,* 1038–1052. doi:10.1037/0022-3514.78.6.1038

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research, 17,* 460–469. doi:10.2307/3150499

Oppenheimer, D. M., LeBoeuf, R. A., & Brewer, N. T. (2008). Anchors aweigh: A demonstration of cross-modality anchoring and magnitude priming. *Cognition, 106,* 13–26. doi:10.1016/j.cognition.2006.12.008

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing, 49,* 41–50. doi:10.2307/1251430

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing, 64,* 12–40.

Parducci, A. (1965). Category judgment: A range–frequency model. *Psychological Review, 72,* 407–418. doi:10.1037/h0022602

Ross, J., & Di Lollo, V. (1968). Category scales and contrast effects with lifted weights. *Journal of Experimental Psychology, 78,* 547–550. doi:10.1037/h0026622

Ross, J., & Di Lollo, V. (1971). Judgment and response in magnitude estimation. *Psychological Review, 78,* 515–527. doi:10.1037/h0031784

Sherif, M., Taub, D., & Hovland, C. I. (1958). Assimilation and contrast effects of anchoring stimuli on judgments. *Journal of Experimental Psychology, 55,* 150–155. doi:10.1037/h0048784

Sherman, S. J., Ahlm, K., Berman, L., & Lynn, S. (1978). Contrast effects and their relationship to subsequent behavior. *Journal of Experimental Social Psychology, 14,* 340–350. doi:10.1016/0022-1031(78)90030-6

Stevens, S. S. (1958). Adaptation-level vs. the relativity of judgment. *American Journal of Psychology, 71,* 633–646. doi:10.2307/1420322

Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology, 73,* 437–446. doi:10.1037/0022-3514.73.3.437

Tversky, A., & Kahneman, D. (1974, September 27). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131. doi:10.1126/science.185.4157.1124

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

Upshaw, H. S. (1984). Output processes in judgment. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 3, pp. 237–256). Hillsdale, NJ: Erlbaum.

Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology: General, 125,* 387–402. doi:10.1037/0096-3445.125.4.387

Wong, K. F. E., & Kwong, J. Y. Y. (2000). Is 7300 m equal to 7.3 km? Same semantics but different anchoring effects. *Organizational Behavior and Human Decision Processes, 82,* 314–333. doi:10.1006/obhd.2000.2900

Woodruff, R. B., Cadotte, E. R., & Jenkins, R. L. (1983). Modeling consumer satisfactions processes using experience-based norms. *Journal of Marketing Research, 20,* 296–304. doi:10.2307/3151833

Received February 18, 2011
Revision received April 4, 2011
Accepted April 5, 2011 ■