Are Adjustments Insufficient?

Nicholas Epley
Harvard University

Thomas Gilovich
Cornell University

Many judgmental biases are thought to be the product of insufficient adjustment from an initial anchor value. Nearly all existing evidence of insufficient adjustment, however, comes from an experimental paradigm that evidence indicates does not involve adjustment at all. In this article, the authors first provide further evidence that some kinds of anchors (those that are self-generated and known to be incorrect but close to the correct answer) activate processes of adjustment, whereas others (uncertain anchors provided by an external source) do not. It is then shown that adjustment from self-generated anchors does indeed tend to be insufficient, both by comparing the estimates of participants starting from different anchor values and by comparing estimates with actual answers. Thus, evidence is provided of adjustment-based anchoring effects similar to the accessibility-based anchoring effects observed in the traditional anchoring paradigm, supporting theories of social judgment that rely on mechanisms of insufficient adjustment.

Keywords: *anchoring; heuristics; biases; insufficient adjustment; judgment and decision making; social judgment*

Everyday life requires us to estimate uncertain qualities: the length of a military campaign, the number of layoffs in a looming recession, or the likely number of lives lost in an act of terrorism. One way to do so is to adjust some salient value to accommodate the particular details of the judgment at hand. Thus, one might start with the average length of past military campaigns, the number of layoffs in the previous recession, or the number of lives lost in a "worst case" scenario.

Judgments rendered in this way conform to what Tversky and Kahneman (1974) described as an anchoring and adjustment heuristic: People anchor on a readily accessible value and adjust from it to estimate the true value. Tversky and Kahneman noted, however, that adjustments tend to be insufficient, with final judgments too close to the initial anchor.

Most of the evidence for insufficient adjustment comes from a laboratory paradigm developed by Tversky and Kahneman (1974) in which participants are first asked to make a comparative assessment (e.g., Was Mohammed born before or after 800 A.D.?) followed by an absolute estimate (e.g., In what year was Mohammed born?). In the most carefully constructed versions of this paradigm, the initial comparison value (800 A.D.) is randomly selected so it is clear to participants that it contains no implication about the true answer. Countless experiments using this paradigm have demonstrated that people's absolute estimates are biased in the direction of the initial comparison value, a finding that has long been considered evidence of insufficient adjustment.

Recent research, however, challenges this interpretation. In particular, it appears that responses in this paradigm are not the product of adjustment at all but rather of semantic priming (Mussweiler & Strack, 1999a, 1999b, 2000; Mussweiler, Strack, & Pfeiffer, 2000; Strack & Mussweiler, 1997). Briefly, it seems that considering the comparative question leads people to assess whether the anchor value might be the correct answer ("Was Mohammed born in 800 A.D.?"). Because people evaluate hypotheses by seeking confirmatory evidence, this

Authors' Note: This research was supported by a junior scholar position to the Peter Wall Institute at the University of British Columbia and National Science Foundation Grant SES0241544 to Epley and National Science Foundation Grants SES0091416 and SES0241638 to Gilovich. We thank Kevin Van Aelst, Andy Heck, Celene Menschel, and Beverly Whelan for help conducting these experiments, Joyce Ehrlinger and John Slemmer for technical assistance with portions of Study 2, and Thomas Mussweiler for helpful comments on a previous version of this manuscript. Correspondence concerning this article should be addressed to Nicholas Epley, Department of Psychology, William James Hall, 33 Kirkland Street, Harvard University, Cambridge, MA 02138; e-mail: epley@wjh.harvard.edu.

PSPB, Vol. 30 No. 4, April 2004 447-460
DOI: 10.1177/0146167203261889

assessment disproportionately activates information consistent with the anchor. This recently activated information, in turn, is highly accessible when the absolute judgment is made, thereby biasing participants' final estimates in the direction of the original anchor.[1] Participants' responses in the standard anchoring paradigm, then, appear to tell us nothing about the nature, or limitations, of mental adjustment.

But this insight is relatively recent, and earlier psychologists working without it found the insufficient adjustment account of the traditional anchoring paradigm useful for explaining a host of everyday judgments, including preference reversals (Lichtenstein & Slovic, 1971; Schkade & Johnson, 1989), probability estimates (Plous, 1989; Wright & Anderson, 1989), hindsight bias (Fischhoff & Beyth, 1975; Hawkins & Hastie, 1991), language production and comprehension (Keysar & Barr, 2002), and various egocentric biases such as the "spotlight effect" and the "illusion of transparency" (Gilovich, Medvec, & Savitsky, 2000; Gilovich, Savitsky, & Medvec, 1998). Quattrone (1982), for example, proposed an anchoring and adjustment account of the correspondence bias—the commonly observed tendency for people to draw overly dispositional inferences about actors laboring under situational constraints. Consistent with this account, manipulations of cognitive load and time pressure that theoretically diminish the ability to engage in effortful adjustment have been found to increase the correspondence bias (for reviews, see Gilbert 1989, 2002), and those who are especially fond of effortful thinking (Caccioppo & Petty, 1982) may not commit the bias at all (D'Agostino & Fincher-Kiefer, 1992). Although these results and others like them are consistent with the insufficiency of adjustment, they are also consistent with other models of human judgment based on improper weighting or integration of information (e.g., Trope, 1986; Trope & Gaunt, 2000). Cognitive load is a coarse experimental manipulation, affecting not only adjustment processes but all kinds of effortful cognition. A significant effect of cognitive load might mean that two or more psychological processes are operating, but this manipulation cannot so clearly inform the interaction between them (Logan, 1997). Perhaps for this reason, researchers after Quattrone have generally stopped calling these "adjustment" processes and have instead preferred the more agnostic term, "correction" processes (e.g., Gilbert, 1989; Wegener & Petty, 1995).

Thus, despite its role as the linchpin of numerous theories of human judgment, there is—astonishingly—no conclusive evidence that adjustment tends to be insufficient. The research reported here seeks to remedy this shortcoming.

We have shown elsewhere that people do engage in serial adjustment outside the standard anchoring paradigm (Epley, in press; Epley & Gilovich, 2001). Specifically, people do in fact adjust from anchor values they generate themselves as "ballpark" figures known to be close to the right answer but in need of modification. How long, for example, does it take Mars to orbit the sun? Most people know that it takes longer than 365 days, but how much longer? Or what is the boiling point of water on Mt. Everest? Most people know it must be less than 212° F (100° C), but how much less? These "self-generated" anchor values serve as starting points in the solution to many everyday judgment problems, and they differ from those in the standard paradigm because they are known to be wrong from the beginning but close to the right answer. There is thus no need to evaluate whether the anchor is correct, and thus no initial (confirmation) hypothesis testing to bias final judgments. Instead, people must adjust this ballpark value until a plausible response is reached.

Consistent with this account, protocol analyses reveal that people verbally articulate a process of anchoring and adjustment when solving judgment problems that elicit self-generated anchors but not when answering questions in the standard anchoring paradigm. In addition, people provide estimates closer to self-generated anchor values (and do so more quickly) when nodding their heads than when shaking their heads, presumably because the former movements make people more likely to accept values early in the adjustment process (Forster & Strack, 1996, 1997; Wells & Petty, 1980). Of key importance, such head movements do not influence responses in the standard anchoring paradigm, providing further evidence that the processes engaged in that paradigm are different from those engaged when individuals themselves generate the starting value that needs adjustment. Finally, putting participants under cognitive load systematically shortens adjustment from self-generated anchors, and inducing effortful thinking through incentives significantly lengthens adjustment (Epley, in press)—manipulations that do not influence responses to either plausible (Chapman & Johnson, 2002; Epley, in press; Tversky & Kahneman, 1974) or implausible (Mussweiler & Strack, 1999b; Wilson, Houston, Etling, & Brekke, 1996) anchors within the standard anchoring paradigm. Not all anchors, it appears, are alike.

This new evidence of serial adjustment from self-generated anchors, coupled with a better understanding of the actual mechanisms involved in the standard anchoring paradigm (and the ambiguities inherent in the "busyness" manipulations used by attribution paradigms), means that one of the most critical features of the anchoring and adjustment heuristic must be tested anew: Do adjustments from self-generated anchors tend to be insufficient? To examine this question, we first sought to provide additional evidence that people do indeed adjust from self-generated anchors (Studies 1 and pretest to Study 4), a judgment process that is not engaged in the standard anchoring paradigm. We then examined whether adjustments from self-generated

anchor values tend to be insufficient by comparing participants' estimates to the actual answers (Studies 3 and 4) and by comparing the estimates of groups of participants who were induced to begin from different anchors (Studies 2 and 4).

STUDY 1

Participants in this experiment were asked two sets of anchoring questions—one that reliably activates self-generated anchor values that are known to be close to the right answer but in need of adjustment (Epley & Gilovich, 2001) and another that has been used in previous research within the standard anchoring paradigm (Jacowitz & Kahneman, 1995). Participants answered these questions either while pulling up on the table at which they were seated (arm flexion) or while pushing down on the table (arm extension).

These hand movements produce distinct evaluative responses. In particular, people tend to evaluate stimuli more favorably when adopting an approach posture through arm flexion than when adopting an avoidance posture through arm extension (Cacciopo, Priester, & Berntson, 1993). We predicted that these different evaluations would influence the magnitude of participants' adjustments. Because adjustment involves movement from an initial anchor until a sufficiently plausible value is reached, the evaluative consequences of arm flexion ought to lead to relatively favorable evaluations of values entertained early in the adjustment process and thereby shorten adjustment. Arm extension, in contrast, ought to lead to relatively unfavorable evaluations of values entertained early on and hence lengthen adjustment. Such a pattern of results would be analogous to the head nodding and shaking effects mentioned earlier and would provide further evidence of serial adjustment from self-generated anchors.

In contrast, we did not expect arm flexion or arm extension to influence responses in the standard anchoring paradigm because, as we have discussed, the anchoring effects observed there appear to be the result of selective accessibility, not (insufficient) adjustment (Mussweiler & Strack, 1999a; Strack & Mussweiler, 1997). Experimenter-provided anchors therefore influence judgment because they alter the *kind* of information that is brought to mind, not the *amount* of information brought to mind in the manner we have just described. Consistent with this analysis, responses in the standard anchoring paradigm are influenced by the compatibility of an anchor and target (Chapman & Johnson, 2002; Strack & Mussweiler, 1997), whether category or exemplar knowledge is activated by an anchor (Mussweiler & Strack, 2000), and by framing manipulations (Mussweiler & Strack, 1999b), all of which are driven by the kind—but not the amount—of information generated while making a judgment. In addition, diminishing the amount of information considered by placing participants under cognitive load does not alter responses in the standard anchoring paradigm, and neither does increasing effort through incentives for accuracy (for reviews, see Chapman & Johnson, 2002; Epley, in press). Because pushing and pulling may influence the amount of information people consider (Friedman & Foerster, 2000, 2002), but not the kind of information generated, we predicted no influence of arm movement on responses in the standard anchoring paradigm.

Method

Thirty-four Harvard University undergraduates participated in exchange for $6. After arriving at the laboratory, participants were seated in front of an Apple G4 computer and asked to don a pair of headphones. They were informed that the experiment was a study of product evaluations and that they would be asked, at the end of the experiment, to evaluate the headphones on a variety of different dimensions. Participants were further informed that we were particularly interested in people's "implicit evaluations" of the headphones—evaluations that are formed "without conscious intention or effort." As a result, participants were told that they would be distracted from thinking about the quality of the headphones by answering a series of questions that would be read to them (over the headphones) and by holding a prescribed body position throughout the experiment. Participants were then randomly assigned to the arm flexion or arm extension condition. Depending on condition, participants were then asked to place both hands either under or on top of the table and to apply mild pressure against the tabletop. Participants were told to hold their prescribed position throughout the questioning period until told by the experimenter to stop.

Participants were asked four self-generated anchoring items—two requiring upward adjustment and two downward adjustment—and four experimenter-provided anchoring items taken from Jacowitz and Kahneman (1995)—two involving high anchor values and two low anchor values.[2] Before answering these items, the experimenter explained that they would be asked two kinds of items, one that involved a single question (i.e., the self-generated items) and another that involved a pair of related questions (i.e., the experimenter-provided items). The experimenter explained that the former involved answering a single general knowledge question and the latter involved an initial comparison with a randomly selected value followed by an absolute estimate. Each anchoring question was presented on the computer screen and simultaneously read through a prerecorded sound file within a Powerpoint presentation. Participants provided their answers verbally, whereupon the experimenter recorded participants' responses and advanced the presentation. The order in which participants answered self-generated and experimenter-provided

anchoring questions was counterbalanced but did not influence any of the results and is not discussed further.

After answering all anchoring questions, participants completed a follow-up questionnaire that assessed whether they knew the intended self-generated anchor values (e.g., the year the United Stated declared its independence) and whether they had thought of that value when generating their estimate. Note that participants in this and the following experiments were not asked to report on the process that led to their judgments. Rather, participants were merely asked whether they knew a particular anchor value and whether it occurred to them while answering the question—mental *contents* that can be reported more accurately than mental *processes* (Nisbett & Wilson, 1977).

Results

To provide a valid test of our hypotheses regarding self-generated anchors, participants needed to (a) know the intended anchor value and (b) have recalled it when making their estimate. Participants who did not meet both criteria were excluded on an item-by-item basis. The number of participants who did not meet these criteria varied across items, but participants met these criteria on an average of 2.91 of the 4 self-generated anchoring items, and all participants met the criteria for at least one question.

To determine whether arm flexion and arm extension influenced participants' responses, we first calculated the absolute difference between participants' estimates and the original anchor value for each item to obtain a measure of adjustment (see also Epley & Gilovich, 2001). These differences were then standardized and averaged separately across the self-generated and experimenter-provided anchoring items such that negative *z* scores indicated responses closer to the original anchor value (i.e., less adjustment). As can be seen in Table 1, participants' arm movements appeared to influence their responses to the self-generated anchoring questions but not the experimenter-provided anchoring questions. For the self-generated anchoring questions, participants engaging in arm flexion provided estimates closer to self-generated anchors than did participants engaging in arm extension, $t(32) = 2.78$, $p < .01$. In contrast, there was no difference across the two arm movement conditions in participants' responses to the experimenter-provided anchoring questions, $t < 1$. A 2 (arm movement: pushing or pulling) × 2 (anchoring item: self-generated or experimenter-provided) ANOVA with repeated measures on the last factor yielded the predicted significant interaction, $F(1, 32) = 8.25$, $p < .01$.

Participants' adjustments from a given self-generated anchor value were all in the same direction (e.g., everyone adjusted upward from 1776 when estimating when George Washington was first elected president). This was not the case with respect to experimenter-provided anchors. A small minority of participants provided responses that were higher than a high anchor value or lower than a low anchor value. To ensure that this did not contribute to the differential effect of arm movements on self-generated versus experimenter-provided anchors, we conducted an additional analysis that excluded all such "contrarian" estimates. Even with these responses excluded, the composite index across the four experimenter-provided anchoring items showed no significant difference between arm flexion ($M = -.01$) and arm extension ($M = -.07$), $t < 1$, and the interaction between arm movement and type of anchoring question remained significant, $F(1, 32) = 4.60$, $p < .05$.

TABLE 1: Standardized Responses to Self-Generated and Experimenter-Provided Anchors Among Participants Engaging in Arm Flexion Versus Arm Extension (Study 1)

	Standardized Estimate	
Anchors	*Arm Flexion*	*Arm Extension*
Self-generated	–.32	.24
Experimenter provided	.07	–.03

Discussion

These results establish a critical precondition for our examination of whether adjustment tends to be insufficient by providing additional evidence that responses to self-generated anchors activate serial adjustment, whereas the anchor values presented in the standard anchoring paradigm do not. A manipulation that past research indicates should influence participants' inclination to view propositions favorably was found to influence responses to self-generated anchoring questions but not responses to the type of questions asked in the standard anchoring paradigm. This suggests that participants were adjusting from self-generated anchor values, testing the adequacy of candidate values removed from the anchor value itself, and terminating adjustment if the candidate value was deemed adequate. Similar results were not obtained with experimenter-provided anchor values. This replicates the results of Epley and Gilovich (2001), but with a rather different manipulation of the likelihood of accepting candidate values considered early in the process of adjustment.

Anchoring effects are routinely obtained in the standard anchoring paradigm, of course, but they appear to be the product of the enhanced accessibility of anchor-consistent information, not of the serial adjustment processes documented in Study 1 for self-generated anchor values. The critical question to examine, then, is whether analogous anchoring effects are reliably obtained when the question or task at hand yields self-generated anchor values. In other words, does true adjustment tend to be insufficient? Although Study 1 was not designed to

TABLE 2: Answers to Self-Generated Anchoring Items Among Participants Engaging in Arm Flexion Versus Arm Extension

		Estimate		
Self-Generated Anchors	*Anchor*	*Answer*	*Arm Flexion*	*Arm Extension*
When was Washington elected President of United States?	1776	1789	1780.58***	1783.07**
What is the boiling point of water on Mt. Everest?	212° F	167° F	189.36° F*	138.86° F
When did the second explorer after Columbus land in the West Indies?	1492	1501	1504.33	1509.83
What is the freezing point of Vodka?	32° F	–20° F	18.48° F***	9.88° F***

NOTE: Estimates that indicate significantly insufficient adjustment (by comparison to the actual answer) are marked by an asterisk.
*$p < .05$. **$p < .01$. ***$p < .001$.

address this question, the data nonetheless certainly support this possibility. Table 2 presents participants' responses to the self-generated questions along with both the initial anchor value and the actual target value they were trying to estimate. These were not uninfluenced estimates, of course. Some participants were engaged in arm flexion, which should have the effect of truncating the process of adjustment relatively quickly, and indeed, their estimates fell significantly short of the actual value on three of the four items. Other participants were engaged in arm extension, which should have the effect of lengthening the process of adjustment. But even the estimates of these participants fell significantly short of the actual value on two of the four items. From these very limited and admittedly contaminated data, it appears that insufficient adjustment may be common. The next three studies were designed to address this issue more systematically.

STUDY 2

Participants in this study were asked to estimate either the number of U.S. states in 1840 or the year in which the Jamestown colony was founded. Both questions were chosen so we could manipulate—without mentioning a particular anchor value—whether participants were likely to anchor on, and adjust from, one of two dates. In the first case, we manipulated whether participants were likely to anchor on the current 50 states and adjust downward or on the original 13 colonies and adjust upward. With respect to the founding of Jamestown, we manipulated whether participants were likely to anchor on 1492 and adjust upward or on 1776 and adjust downward. We predicted that participants induced to generate different anchor values would arrive at significantly—and predictably—different estimates. We also predicted that participants' estimates would fall short of the actual answers to these questions in a manner consistent with insufficient adjustment. We predicted that participants would underestimate the 26 states in 1840 if they were led to adjust from the nation's original 13 colonies but overestimate that value if they were led to adjust from the nation's current 50 states. Similarly, we predicted that participants would estimate that Jamestown was founded earlier than the actual date of 1607 if they were induced to anchor on 1492 but estimate that it was founded later than 1607 if they were induced to anchor on 1776.

Method

One group of 190 Cornell University undergraduates was asked as part of a larger internet survey to estimate the number of states in the United States in 1840. Participants in one condition did so after reading that "the United States declared its independence on July 4, 1776," a statement designed to prime their knowledge of the 13 original colonies and to induce them to adjust from that value. Participants in the other condition confronted the key question after reading that "the United States will celebrate its 225th anniversary on July 4, 2001." This was designed to prime their knowledge of the current 50 states and induce them to adjust from that value. Due to a programming mistake, approximately one quarter of the participants were randomly assigned to the first condition ($n = 49$) and the remaining to the second ($n = 141$).

A separate group of 88 Cornell undergraduates was asked, in a classroom setting, to estimate when the Jamestown Colony was established. In this case, the attempt to influence the participant's self-generated anchor was embedded in the question itself. One group was asked, "In what year do you think Jamestown, the first English settlement in the New World, was established?" The use of "New World" was designed to trigger an association to the Spanish explorers and 1492. The other group was asked the same question, except that "New World" was replaced by "United States" in an effort to trigger an association to the nation's founding and 1776.

After making their estimates, participants in both versions were asked if they had thought of any other number(s) when generating their answers. Participants who estimated the number of U.S. states in 1840 were simply asked, in an open-ended format, whether they had recalled any other number when generating their estimate. Those who estimated the date of the founding of Jamestown were asked specifically if they had thought of 1492 (yes or no) when generating their estimate and whether they had thought of 1776 (yes or no).

Results

Not surprisingly, participants were more likely to state that they began their process of judgment from the intended anchor value when they were asked about specific dates (74.4% in the Jamestown version) than when they were simply asked if they had considered "any number" (46% in the U.S. states version). If participants in the latter version started with an initial value and adjusted it to reach a final estimate, they would have considered several different numbers, making our question difficult to answer. Whatever the cause, all initial analyses included all participants in the sample. Subsequent analyses were based only on the responses of those who reported considering one of the intended anchor values.

Between-condition analyses. As predicted, participants induced to consider different anchor values arrived at different final estimates. Those induced to anchor on the original 13 colonies estimated that there were fewer states in 1840 (M = 21.3) than did those induced to anchor on America's current 50 states (M = 30.9), $t(188) = 4.86$, $p < .001$. Similarly, those induced to anchor on 1492 estimated that Jamestown was founded earlier (M = 1601.5) than those induced to anchor on 1776 (M = 1641.1), $t(86) = 2.57$, $p < .05$. Those who began the process of adjustment from different anchors ended up with predictably divergent answers, indicating that adjustment—true adjustment from self-generated anchor values—tends to be insufficient.

Internal analyses. If this account is valid, then the between-condition effects should only get stronger if we restrict the analysis to those participants, regardless of condition, who reported anchoring on either of the two pertinent values. Note that the between-condition effects were obtained because our manipulations were effective in leading participants in the different conditions to anchor on different values. In particular, of those who reported having entertained a value other than their final estimate, 81% reported having considered 13 after being primed with the Declaration of Independence but only 38% did so after being primed with America's 225th anniversary. Conversely, 62% reported having considered the value 50 after being primed with America's 225th anniversary, but only 19% did so after being primed with the Declaration of Independence. This pattern was statistically significant, $\chi^2(1) = 13.07$, $p < .01$. For the Jamestown question, of those who reported having entertained a value other than their final estimate, 71% reported having considered the year 1492 after being exposed to the phrase "New World" but only 9% did so after being exposed to the phrase "United States." Conversely, 91% reported having considered the year 1776 after being exposed to "United States" but only 29% did so after being exposed to "New World." This difference was also statistically significant, $\chi^2(1) = 27.20$, $p < .001$.

Ignoring experimental condition, participants who reported anchoring on 13 states estimated that there were significantly fewer states in 1840 (M = 20.3) than did those who reported anchoring on 50 (M = 39.0), $t(84) = 10.73$, $p < .001$. Similarly, participants who reported having considered 1492 when generating their estimate of the date of Jamestown's founding thought that it occurred earlier (M = 1547.0) than did those who reported having considered 1776 (M = 1673.7), $t(64) = 8.63$, $p < .001$.

Comparing estimates to reality. As mentioned previously, another way to assess whether adjustments tend to be insufficient is to examine whether participants' estimates differ not from one another but from the actual answer. Did participants' adjustments make it all the way to the actual answer? The results indicate that they did not.

Recall that there were 26 states in 1840. When the data from all participants were examined, regardless of whether they articulated having entertained the intended self-generated anchor value, those primed with the Declaration of Independence provided estimates that were significantly lower than this value, $t(48) = -4.08$, $p < .001$. Those primed with the U.S.'s 225th anniversary, in contrast, provided estimates that were significantly higher than this value, $t(140) = 4.48$, $p < .001$. In addition, the estimates of those who reported having considered the original 13 states significantly underestimated the 26 states in 1840, $t(43) = -6.5$, $p < .0001$, and those who reported having considered the current 50 states significantly overestimated the actual number of states in 1840, $t(41) = 8.51$, $p < .0001$.

Recall also that Jamestown was founded in 1607. When the data from all participants who were asked to estimate this value were examined, regardless of whether they reported having considered the intended self-generated anchor value, those primed with the "New World" (and hence 1492) estimated that Jamestown was founded earlier than it actually was, although not significantly so, $t < 1$. Those primed with the "United States" (and hence 1776), in contrast, estimated that Jamestown was founded later than it actually was, $t(35) = 4.02$, $p < .001$. In addition, those who reported having considered 1492 significantly underestimated the founding date of 1607, $t(25) = 6.19$, $p < .0001$, and those who explicitly mentioned having considered the year 1776 significantly overestimated the actual founding date, $t(40) = 6.79$, $p < .001$.

Discussion

The results of this study provide clear evidence of insufficient adjustment. Those led to consider different anchor values provided significantly different estimates. Those induced to estimate the number of U.S. states in

1840 by adjusting upward from the original 13 colonies provided lower estimates than those induced to adjust downward from the current 50 states. Likewise, those induced to estimate the date of the founding of Jamestown by adjusting upward from 1492 provided lower estimates than those induced to adjust downward from 1776.

Furthermore, the adjustments made from these anchor values tended to fall short of the actual values participants were trying to estimate. Those adjusting upward from the 13 colonies significantly underestimated the actual number of states in 1840, and those adjusting downward from the current 50 states significantly overestimated the actual number of states. Likewise, those adjusting upward from 1492 tended to estimate that Jamestown was founded earlier than it actually was, and those adjusting downward from 1776 estimated that it was founded later than it actually was.

Note, however, that this latter evidence of insufficient adjustment—predictable deviation from the actual answer—is based only on four comparisons involving two different content items. One might reasonably question the generality of this result. Study 3 was conducted to address this issue.

Note also that we did not explicitly demonstrate that participants did, in fact, engage in true serial adjustment from the induced anchor values. Instead, we relied on previous findings, that is, the results of Study 1 and Epley and Gilovich (2001), indicating that people do indeed adjust from self-generated anchors. To dispel even more strongly this concern that participants may not have engaged in true adjustment, Study 3 used many of the exact questions from past research for which adjustment has been clearly established.

STUDY 3

Participants in this study answered 10 questions that have been shown previously to invoke serial adjustment (Study 1; Epley & Gilovich, 2001). Based on the results of Study 2, we hypothesized that adjustments would be insufficient. In this study, however, we assessed the insufficiency of adjustment by comparing participants' estimates with the actual values in question. We predicted that participants' estimates would, on average, fall in between their initial self-generated anchor values and the actual answers.

One difficulty in using this procedure to assess the insufficiency of adjustment is that it is not clear how to sample from the population of situations in which people engage in anchoring and adjustment. Thus, any demonstration of insufficient adjustment might result from some peculiarity of the particular sample of questions used. Two elements of our procedure dampen this concern.

First, to construct a question that will elicit demonstrable adjustment from self-generated anchors, it is essential that (a) very few participants know the true answer and thus must estimate it and (b) nearly all participants will spontaneously think of the same value as a starting point for their estimates. We have found in past research, however, that there is little truly common knowledge, even among ivy-league university participants, and thus few questions that meet this latter criterion. We thought initially that questions involving such seemingly memorable values as 714 (Babe Ruth's career home run total) or 28 (days required for the moon to orbit the earth) would satisfy the common-knowledge criterion. They did not. Common knowledge appears to be less common than one might expect. The questions used in Study 3, then, are those that survived a great deal of pretesting involving this common-knowledge test and involve such basic information as the boiling point of water at sea level (212° F), the duration of the earth's orbit (365 days), the number of states in the United States (50), the date of the "discovery" of the New World (1492), and the freezing point of water (32° F). Although more clever minds than ours could doubtless generate additional items that meet this criterion, the key point is that the sample of items used in Study 3 is not a set that has been selected because of the likelihood of eliciting evidence of insufficient adjustment. Rather, it represents our best efforts to examine the full population of useable items.

Second, we sought to diminish any concern about this particular set of questions by manipulating the amount participants needed to adjust from the same anchor value. For example, some participants were asked how long it takes Mars to orbit the sun, whereas others were asked how long it takes Neptune to orbit the sun. If anchoring and adjustment is a useful but imperfect heuristic, then estimates should be responsive to the actual amount of adjustment required but generally fall short of the actual answer. We were particularly interested in whether participants' estimates would be insufficient even when very little serial adjustment was required. Whenever possible, then, we asked participants to estimate a value that was the closest to the self-generated anchor and thus required the least adjustment. Participants therefore estimated the orbit of the next furthest planet from Earth (i.e., the next longest to 365 days), the landing in the West Indies of the second European explorer (i.e., the next landing after 1492), and the start of the second U.S. presidency (John Adams—the next after Washington's). These "nearest neighbor" questions were included to reduce the concern that adjustments tend to be insufficient only when there happens to be a considerable distance from the anchor to the target value.

Method

Seventy-eight Cornell University undergraduates were asked 10 self-generated anchoring questions

adapted from Epley and Gilovich (2001). Seven of these items contained a between-participants manipulation of the question's target (see Table 3). Approximately half of the participants estimated values requiring relatively little adjustment from the self-generated anchor (e.g., How long does it take the planet Mars to orbit the Sun?), whereas the others considered targets that required relatively more adjustment (e.g., How long does it take the planet Neptune to orbit the Sun?). An additional three items that did not lend themselves to this between-participants manipulation were asked of all participants to broaden our sample.

After answering the 10 questions, participants completed a questionnaire that assessed whether they knew the intended anchor value (e.g., "How many days does it take the earth to orbit the sun?") and whether they had thought of that value when generating their answer.

Results

Three conditions had to be met for an adequate test of our hypotheses about self-generated anchors (Epley & Gilovich, 2001). First, participants had to know the intended anchor value. Second, they had to report having thought of that value when making their estimate. Finally, because we were interested in the sufficiency of adjustment, participants needed to adjust in the intended direction.[3] Participants who did not meet these conditions were excluded on an item-by-item basis. This left an average of 80% of the original sample for a given item, ranging from as few as 50% for the death of Jesus' apostles to as many as 99% for the highest body temperature ever recorded in a living human being. Two participants knew none of the intended anchor values, leaving 76 participants in the final analyses.

The data reported in Table 3 clearly demonstrate that participants were sensitive to the amount of adjustment required by each of the questions. For all but one item (the explorers who followed Columbus's maiden voyage to the New World), participants provided estimates indicative of greater adjustment when the question called for a value that, in fact, required greater adjustment. For instance, participants estimated, on average, that Mars takes 574 days to orbit the sun but that Neptune takes 3,447 days. To assess the statistical significance of this pattern, responses to these 14 questions (two versions of each of seven questions) were standardized, reverse scored where appropriate, and averaged into a composite index. Results indicated that participants did indeed adjust less for targets requiring less adjustment ($M = -.22$) than for targets requiring more adjustment ($M = .28$), $t(74) = 4.86$, $p < .001$.

To determine whether these adjustments were sufficient, participants' estimates for each of the 17 questions (7 involving the between-participants manipulation plus 3 others) were compared to the actual answers to these questions. Our assessment of whether participants tended to adjust insufficiently is based on how their mean responses to the 17 items compared to the actual values being estimated. Recall that in addition to the 7 pairs of questions for which the target value was manipulated between subjects, all participants estimated the same target values for an additional three questions. Table 4 presents participants' mean estimates and the actual values for all 17 questions. These data reveal a strong tendency for insufficient adjustment. Of the 17 estimates, the mean value for 14 fell short of the actual answer (binomial $p < .05$). To examine this pattern in more detail, one-sample t tests were performed on each item, comparing participants' estimates with the actual answer. Ten of these t tests revealed significant evidence of insufficient adjustment, and only 2 (the boiling point of water in Denver and the arrival of the fifth explorer after Columbus) revealed significant overadjustment. Finally, these t values were treated as an index variable and submitted to a one-sample t test to determine whether they were, on average, significantly less than a null value of 0. They were ($M = -7.87$), $t(16) = -2.65$, $p < .05$, indicating that adjustments from these self-generated anchors tended to be insufficient.

TABLE 3: Mean Estimates for Self-Generated Anchoring Questions Requiring Large and Small Adjustments (Study 3)

	Adjustment Required		
Question	*Anchor*	*Small*	*Large*
Apostle's year of death (first, last)	33	43.6	58.4
Explorer after Columbus (second, fifth)	1492	1516.0	1514.3
Elected president of United States (Washington, Adams)	1776	1779.8	1793.1
Days to orbit the sun (Mars, Neptune)	365	574.4	3,447.0
Boiling point of water (Denver, Mt. Everest)[a]	212	191.3	176.2
Freezing point (vodka, pure alcohol)[a]	32	1.8	–8.0
States in United States (1880, 1840)[a]	50	39.21	34.21

NOTE: The target of judgment is listed within the parentheses for each item. The target requiring relatively small adjustment is on the left and the target requiring large adjustment is on the right.
a. Items requiring downward adjustment.

Discussion

These data reinforce those obtained in Study 2 and provide converging evidence that adjustments tend to be insufficient. An unbiased set of questions designed to elicit adjustment from self-generated anchor values produced estimates that tended to fall between the anchor value and the target being estimated, indicating that the adjustments were generally insufficient. Thus, whether the sufficiency of adjustment is assessed by comparing the estimates of participants induced to generate differ-

TABLE 4: Mean Estimated and Actual Answer to Self-Generated Anchoring Questions (Study 3)

Question	*Estimated Answer*	*Actual Answer*	t
Orbit of Mars around the Sun	574.4	868.98	–4.98***
Orbit of Neptune around the Sun	3,447.0	60,225	–39.56***
Washington elected president	1779.8	1788	–8.52***
Adams elected president	1793.1	1796	–1.45
Number of states in 1880[a]	39.2	38	–0.79
Number of states in 1840[a]	34.2	26	–4.23**
Death of Jesus' first apostle	43.6	44	–0.10
Death of Jesus' last apostle	58.4	70	–2.8*
Second explorer after Columbus	1516.0	1501	1.64
Fifth explorer after Columbus	1514.3	1505	2.1*
Freezing point of vodka[a]	1.8	–20.00	–5.3***
Freezing point of pure alcohol[a]	–8.0	–174.50	–33.89***
Boiling point of water on Mt. Everest[a]	176.2	167.00	–1.65
Boiling point of water in Denver[a]	191.3	203	3.22**
Highest body temperature ever recorded	109.6	115.70	–11.64***
Lowest body temperature ever recorded [d]	84.9	57.50	–19.91***
Gestation period of elephant	16.8	22	–5.96***
Median			–4.23

NOTE: All temperatures are reported in degrees Fahrenheit. The orbits of Mars and Neptune around the Sun are reported in days. Positive *t* values indicate overly sufficient adjustment (the answer is not between the anchor value and the actual answer). Negative *t* values indicate insufficient adjustment (the answer is between the anchor value and the actual answer).

a. Items requiring downward adjustment.

*$p < .05$. **$p < .01$. ***$p < .001$.

ent anchors (Study 2) or by comparing participants' estimates with the actual answers (Studies 2 and 3), it is clear that adjustments tend to be insufficient.

Although these results suggest that insufficient adjustment is pervasive (it emerged on 14 of the 17 items used in Study 3), it is not universal. There are occasional exceptions, and they are most likely to appear, all else being equal, the closer the target being estimated is from the anchor a person generates. Someone who estimates the inauguration date of President Tyler by adjusting from the inauguration date of his predecessor, William Harrison, would be hard-pressed to underestimate the date in question simply because Harrison's death so soon after taking office leaves little room for estimates to fall short of the target date.

With this consideration in mind, we designed Study 4 to examine further the issue of when adjustments are likely to be insufficient and when they are not. A pair of considerations influenced the design of this study. First, we assume that people are roughly sensitive to the amount of adjustment required from a self-generated anchor value—a finding demonstrated in Study 3. Second, we assume that adjustments are more likely to be insufficient the farther the value to be estimated is from the anchor value they generate. Both of these assumptions are assessed in Study 4.

STUDY 4

Participants in this experiment estimated how many states there were in the United States at one of eight points in history. Because there were no states until 1789 and a full complement of 50 in 1959, participants estimated the pertinent number at one of eight dates in 20-year intervals from 1800 to 1940. Note that participants (as in Study 2) could arrive at their estimates by adjusting from one of two likely anchor values—the original 13 colonies or the current 50 states.

We predicted that participants would be more likely to anchor on 13 when estimating dates closer to 1776 and to anchor on 50 when estimating dates closer to today. Regardless of the anchor used, however, we expected participants to adjust insufficiently and thus provide answers falling between the selected anchor and actual answer. We expected that adjustments would be sensitive to the actual number of states at any given time (that people would estimate more states for later dates) but that those who anchored on 13 would generally underestimate, and those who anchored on 50 would generally overestimate, the actual number at any specified time.

Pretest

Because we relied in this study on participants' self-report of the anchor value they considered instead of manipulating it as we did in Study 2, we thought it prudent to conduct a preliminary investigation to demonstrate conclusively that people utilize the anchoring and adjustment heuristic when estimating the number of U.S. states. To do so, we returned to the priming methodology employed in Study 2, combined with the arm flexion and extension manipulation used in Study 1. Specifically, 63 Harvard university undergraduates estimated how many states were in the United States in 1840. Approximately half were primed to anchor on the original 13 colonies and adjust up, whereas the others were primed to anchor on 50 and adjust down, using the same priming manipulation employed in Study 2. In addition, approximately half answered this question while pulling up with both hands on the table in front of them while the other half did so while pushing down. After estimating the number of states, participants were asked whether they first thought of the number 13 when answering the question (yes or no) and also whether they first thought of the number 50 (yes or no).

As in Study 1, we expected that arm flexion and extension would influence participants' responses in a manner consistent with use of the anchoring and adjustment heuristic. In particular, we predicted that arm flexion would lead people to evaluate possible estimates encountered early in the adjustment process more favor-

ably (Caccioppo et al., 1993) and terminate their information search more quickly (Martin, Ward, Achee, & Wyer, 1993), thereby leading to less adjustment. We predicted that arm extension, in contrast, would lead people to evaluate possible estimates encountered early in the adjustment process less favorably, thereby leading to more adjustment. If participants primed to anchor on 13 are estimating the number of states by serially adjusting upward from that value, they should stop sooner and provide smaller estimates when engaging in arm flexion than when engaging in arm extension. In contrast, if participants primed to anchor on 50 are estimating the number of states by adjusting downward from that value, then these participants should stop sooner and provide higher estimates while engaging in arm flexion than when engaging in arm extension. Figure 1 shows that this is exactly what participants did. Those primed to anchor on 13 provided lower estimates if engaged in arm flexion than if engaged in arm extension, $t(29) = 2.12$, $p < .05$, whereas those primed to anchor on 50 provided higher estimates if engaged in arm flexion than if engaged in arm extension, $t(31) = 2.79$, $p < .01$. A 2 (anchor: 13 or 50) × 2 (arm movement: flexion or extension) ANOVA revealed a significant main effect for anchor, $F(1, 63) = 10.61$, $p < .01$, qualified by the predicted significant interaction, $F(1, 54) = 11.96$, $p < .01$.

Notice that these results also demonstrate that the insufficiency of adjustment can be diminished, and in this case eliminated, by manipulations that increase adjustment. Recall that there were actually 26 states in the Untied States in 1840. Participants adjusted insufficiently from both 13 and 50 when engaged in arm flexion, ts(10 and 12) = 3.74 and 2.95, respectively, both ps < .05, but adjusted sufficiently when engaged in arm extension, both ts < 1.

As in Study 2, the priming manipulation successfully induced participants in the two priming conditions to adjust from different anchor values. Participants primed to anchor on the original 13 colonies were more likely to report having thought of this value (71%) when making their estimate than were participants primed to anchor on 50 states (29%), whereas participants primed to anchor on 50 were more likely to report having thought of that value (67%) than were participants primed to anchor on 13 (33%). This pattern was statistically significant, $\chi^2 = 8.20$, $p < .05$.

Also as in Study 2, the evidence for serial adjustment is even stronger in an analysis based on participants' self-reported anchor values. As shown in Figure 2, participants who reported anchoring on 13 thought there were fewer states in 1840 if they were engaged in arm flexion than if engaged in arm extension, $t(27) = 2.60$, $p < .05$, whereas participants who reported anchoring on 50 thought there were more states in 1840 if engaged in arm flexion than if engaged in arm extension, $t(24) = 2.77$, $p < .05$. A 2 (anchor: 13 or 50) × 2 (arm movement: flexion or extension) ANOVA revealed a main effect for anchor, $F(1, 54) = 65.42$, $p < .01$, qualified by the predicted interaction, $F(1, 54) = 14.41$, $p < .01$.

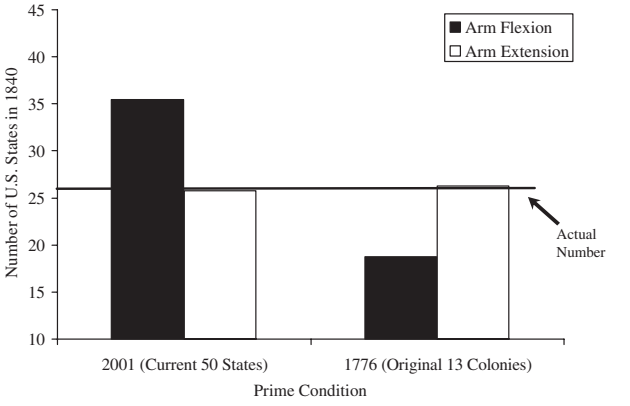

Figure 1 Estimated number of U.S. states in 1840 by participants primed to anchor on 13 or 50 while engaging in arm flexion or extension (pretest, Study 4).

These results demonstrate that anchoring and adjustment underlies our participants' estimates of the number of states in the United States. They also reinforce the validity of participants' self-reports of their initial anchor value. Participants primed to anchor on 13 reported anchoring on that value to a much greater extent than participants primed to anchor on 50, and vice versa. Although self-reports of mental processes are of questionable validity, reports of mental content—such as whether one recalled a give value—are not (Nisbett & Wilson, 1977). We therefore rely in the next experiment on participants' reports of the anchor value they recalled to test whether adjustments across a wide range of targets tend to be insufficient.

Method

Four hundred and seventeen Cornell University undergraduates estimated the number of states that comprised the United States at one of eight points in history: 1800, 1820, 1840, 1860, 1880, 1900, 1920, or 1940. Immediately after making this estimate, participants were asked if they had thought of "the number 13 when making [their] estimate" (yes or no) and if they had thought of "the number 50 when making [their] estimate" (yes or no).

Results and Discussion

Only participants who reported considering an anchor of either 13 or 50 when making their estimate were retained in the following analysis (89.3% of the total sample).

As can be seen in Table 5, participants were more likely to report having anchored on 13 when estimating the number of U.S. states early in the country's history than when estimating the number of states later in the

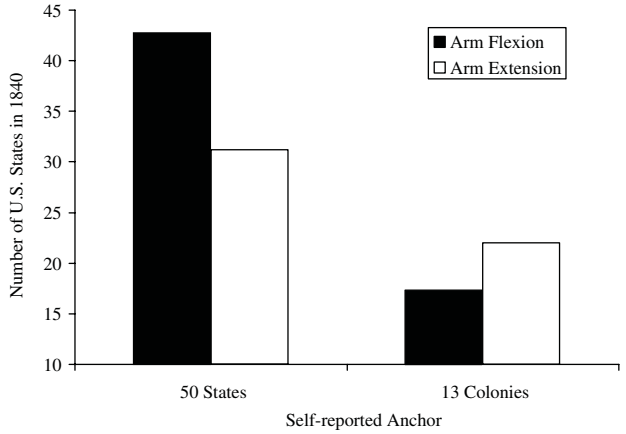

Figure 2 **Estimated number of U.S. states in 1840 by participants who reported anchoring on 13 and participants who reported anchoring on 50 while engaging in arm flexion or extension (pretest, Study 4).**

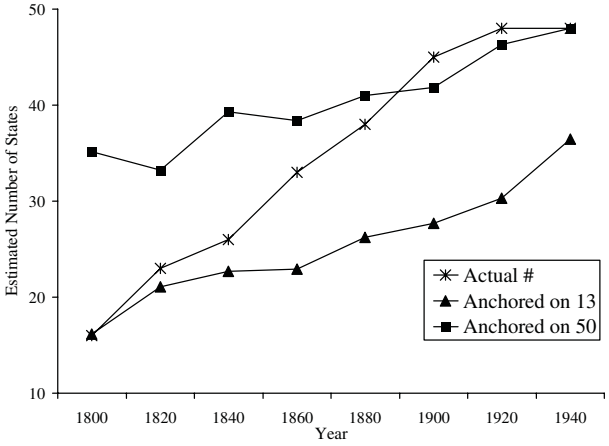

Figure 3 **Number of states that comprised the United States at eight points in time and estimates of the number of states at those times made by participants who reported anchoring on either 13 or 50 states (Study 4).**

TABLE 5: Percentage of Valid Participants Who Anchored on 13 and 50 for Each of the Eight Dates Considered From 1800 to 1940 (Study 4)

Year	*Anchored on 13*	*Anchored on 50*
1800	85	15
1820	68	32
1840	59	41
1860	50	50
1880	58	42
1900	14	86
1920	14	86
1940	4	96

country's history. This suggests that participants tended to anchor on a value they knew to be close to the right answer, but not quite right. In general, estimates were sensitive to the actual number of states in the United States at a given time, with lower estimates for earlier than for later dates, $F(7, 371) = 8.81$, $p < .001$ (see Figure 3).

Participants who anchored on 13 consistently gave lower estimates than those who anchored on 50, $F(1, 371) = 155.51$, $p < .001$, suggesting that adjustments tended to be insufficient. Further evidence of insufficient adjustment was obtained by comparing participants' estimates for a given year with the actual number of U.S. states at that time. Specifically, each participant's estimate was subtracted from the actual number of states in that particular year. On average, participants who anchored on 13 significantly underestimated ($M = 5.79$), $t(159) = -7.26$, $p < .001$, and those who anchored on 50 significantly overestimated ($M = -2.33$), $t(211) = 3.55$, $p < .001$, the actual number of states. The extent to which adjustment was insufficient, however, was qualified by the year considered, $F(7, 371) = 21.07$, $p < .001$. When the actual number of states in a particular year was close to the selected anchor value, adjustments were generally sufficient; when the actual number of states was quite different from the selected anchor value, adjustments were substantially insufficient.

These findings are consistent with those of Study 3. Estimates were tied to the actual answers but nevertheless were consistently biased in the direction of the original anchor value. These results cannot be due to a biased sample of targets because targets across the entire range were selected at systematic intervals. Insufficient adjustment appears to be common and often pronounced.

GENERAL DISCUSSION

These results clearly demonstrate that the adjustments triggered by the anchoring and adjustment heuristic tend to be insufficient. The estimates of participants induced to adjust from different anchor values fell short of one another (Study 2) and also tended to fall short of the actual values being estimated (Studies 2, 3, and 4). These results are of critical importance because insufficient adjustment has been used to account for a great many findings in the intuitive judgment and social psychological literatures (Gilbert, 2002; Gilovich & Savitsky, 1999; Hawkins & Hastie, 1991; Keysar & Barr, 2002; Lichtenstein & Slovic, 1971; Rottenstreich & Tversky, 1997).

Important or not, one might have the feeling that all of this is old hat. Didn't the literature on the anchoring and adjustment heuristic establish long ago that adjustments tend to be insufficient? And didn't research on

the correspondence bias, with all those "busyness" manipulations, likewise demonstrate the insufficiency of adjustment? They did not. Surprisingly, perhaps, virtually all existing evidence for insufficient adjustment comes from an experimental paradigm that recent research quite clearly demonstrates does not involve adjustment at all. The biased judgments observed in that paradigm are the result of semantic priming, not insufficient adjustment (Mussweiler et al., 2000; Mussweiler & Strack, 1999a, 1999b, 2000; Strack & Mussweiler, 1997). And, as we discussed earlier, there is some ambiguity as to exactly what the busyness manipulations used in the attribution literature establish (Trope, 1986; Trope & Gaunt, 2000). Furthermore, the adjustments involved when making dispositional inferences may be a rather different sort than that examined here. In the sort of judgments examined here and elsewhere in the anchoring and adjustment literature, the anchor lies on the same numerical scale as the value being estimated. One tries to estimate Mars's 869-day orbit by adjusting from Earth's 365-day orbit. In contrast, when it comes to making dispositional inferences, the "anchor" value is not as precise ("How pro-Castro did that speech seem?"), and it is unclear that it is in the same units as the ultimate judgment ("How pro-Castro is that person?"). Because the processes involved in these two different contexts may very well be quite distinct, evidence of insufficient "correction" in one cannot confidently be taken as evidence of insufficient adjustment in the other.

Although researchers have marshaled considerable evidence that the anchoring effects observed in the standard anchoring paradigm are the result of enhanced accessibility of anchor-consistent information rather than sequential adjustment, Mussweiler and Strack (2000) argue that the processing of "implausible" anchor values (e.g., Did Mahatma Gandhi live to be 271 years old?) does involve some kind of adjustment. Specifically, they propose something of a "one-step" adjustment process whereby people first adjust the implausible value to some outer limit of conceivable values (say, 110 years old). Other processes, specifically those elicited in response to questions involving plausible anchor values, then kick in to yield the final estimate.

Because anchoring effects have been observed in response to implausible anchor values, one might take this as evidence of insufficient adjustment. But such a conclusion is unwarranted. If Mussweiler and Strack are correct, the adjustment in such cases is not intended to reach the correct answer (as in the case of self-generated anchors), only an outer limit of possible values. Any bias in people's final judgments is thus more appropriately attributable to what is done next, to how this "adjusted" anchor value is processed, not to the initial adjustment itself. Furthermore, the strongest existing evidence of adjustment from implausible anchors is a null effect—that differentially extreme anchor values do not produce different anchoring effects (Mussweiler & Strack, 2000). But this is countered by evidence that implausible anchors are not influenced by manipulations of attentional resources, contrary to what one would expect from a deliberate adjustment process, and to what we have observed with self-generated anchors. Finally, Mussweiler and Strack's adjustment account of implausible anchor values suggests that implausible anchors should produce anchoring effects either larger or equivalent to plausible anchor values. Although this pattern has indeed been observed (Chapman & Johnson, 1994; Mussweiler & Strack, 1999a), the opposite also has been reported such that implausible anchors in three experiments actually produced weaker anchoring effects (Wegener, Petty, Detweiler-Bedell, & Jarvis, 2001). It thus appears that implausible anchors are processed differently than self-generated anchors that are known to be close to the right answer, but wrong. Exactly how they are processed differently is a matter for future research.

More easily answered is why the adjustments made from self-generated anchors tend to be insufficient. We believe there are three likely reasons. First, the process of adjustment likely involves some degree of "satisficing." With respect to the self-generated anchors we have investigated, people retrieve a value that is close to the actual value and then adjust to find an acceptable answer. In the absence of a strong motivation for accuracy, people likely accept one of the first acceptable values they encounter during the process of adjustment (Quattrone, Lawrence, Finkel, & Andrus, 1981). Adjustment tends to be insufficient, then, because it stops just as participants reach the edge of their range of plausible values.

Second, the mental operations involved in serial adjustment require conscious attention, and anything that makes a person distracted, hurried, or unwilling to devote attentional resources will systematically shorten adjustment (Gilbert, 1989). Because life is lived in societies rather than vacuums, the many demands of the social world are more likely to distract attention than to focus it, and thus frequently shorten the process of adjustment.

Finally, adjustments may often be insufficient because people prefer to make one type of error (a cautious one) than another (a rash one). If the wrong bus stop were taken, would it be preferable to take the one just after or just before the right one? When estimating the length of the next bull market, is it better to overshoot or undershoot? The answers to such questions, of course, depend on the particular context involved, and there is nothing in the current anchoring literature indicating that people are more willing to make one sort of error than another. But the psychology involved is not dissimilar to that involved in people's thoughts about errors of omission and commission, and there it is clear that people

tend to kick themselves more over (rash) actions than (cautious) inactions (Gilovich & Medvec, 1995; Kahneman & Tversky, 1982; Miller & Taylor, 1995). Whether a similar pattern of regret applies to insufficient and overly sufficient adjustment is currently unknown.

These three processes may explain why insufficient adjustment is the rule and overly sufficient adjustment the exception. Note, however, that we encountered a few such exceptions in our experiments, and we would generally expect to find sufficient or overly sufficient adjustment whenever two conditions are met. First, overly sufficient adjustment is likely to occur whenever the actual answer happens to be close, in absolute terms, to the initial anchor. There is little room for an insufficient adjustment, for example, when estimating the boiling point of water in Denver (203° F) by adjusting down from the boiling point at sea level (212°F). Second, serial adjustment likely consists of a series of "jumps" that people take until they reach a value that seems plausible, and so overly sufficient adjustment is likely to occur whenever the preexisting theories and generic knowledge that guide the size of these jumps is miscalibrated. For example, our participants adjusted too much when estimating the arrival of the second explorer to the West Indies, most likely because they focused on what would seem to be the long time interval involved in financing and sailing small wooden boats across the Atlantic, and not on Spain's interval-shrinking lust for gold and spices. Our experiments manipulated the amount of adjustment within items but did not address the mechanisms that govern insufficient, and occasionally overly sufficient, adjustment between items. A full understanding of these mechanisms must await the outcome of future research.

The tendency to satisfice and the attention-demanding nature of serial adjustment suggests that insufficient adjustment might be overcome by more effortful and careful thought. Indeed, we report elsewhere that monetary incentives for accuracy increase adjustments from self-generated anchors but have no influence on experimenter-provided anchors in the standard anchoring paradigm (Epley, in press). More generally, the attempt to find ways to eliminate or diminish anchoring effects must be guided by knowledge of its varied underlying mechanisms. Interventions that are effective in overcoming the semantic priming effects observed in the standard anchoring paradigm are likely to be quite different from those that succeed in rectifying the serial (and insufficient) adjustment processes we have examined here.

NOTES

1. Not all investigators agree with this semantic priming account of the traditional anchoring paradigm, and some suggest that simple numerical priming (Wilson, Houston, Etling, & Brekke, 1996; Wong & Kwong, 2000) or persuasion (Wegener, Petty, Detweiler-Bedell, & Jarvis, 2001) may account for the existing data. We do not intend to resolve this debate here but wish only to point out a consensus opinion that responses to plausible anchors within the standard anchoring paradigm are not produced by insufficient adjustment (Chapman & Johnson, 2002; Epley, in press).

2. The two self-generated anchors that required upward adjustment (with anchor values in parentheses) were "In what year was George Washington elected President of the United States (1776)?" and "In what year did the second European explorer after Columbus arrive in the West Indies (1492)?" The two self-generated anchors that required downward adjustment were "What is the boiling point of water on Mount Everest (212°F)?" and "What is the freezing point of Vodka (32°F)?" The two experimenter-provided anchoring items involving low anchor values were "Is the population of Chicago more or less than 200,000?" and "Is the tallest Redwood tree more or less than 65 feet tall?" The two experimenter-provided anchoring items involving high anchor values were "Is Mount Everest taller or shorter than 45,000 feet?" and "Is the Mississippi River longer or shorter than 2,000 miles?"

3. This restriction was not required in the other experiments reported in this article because all participants in those experiments adjusted in the appropriate direction.

REFERENCES

Caccioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*, 116-131.

Caccioppo, J. T., Priester, J. R., & Berntson, G. G. (1993). Rudimentary determinants of attitudes. II: Arm flexion and extension have differential effects on attitudes. *Journal of Personality and Social Psychology, 65*, 5-17.

Chapman, G. B., & Johnson, E. J. (1994). The limits of anchoring. *Journal of Behavioral Decision Making, 7*, 223-242.

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). Cambridge, UK: Cambridge University Press.

D'Agostino, P. R., & Fincher-Kiefer, R. (1992). Need for cognition and the correspondence bias. *Social Cognition, 10*, 151-163.

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.

Epley, N., & Gilovich, T. (2003). *Anchoring and insufficient adjustment: When, why, and how to ameliorate.* Manuscript submitted for publication.

Fischhoff, B., & Beyth, R. (1975). "I knew it would happen": Remembered probabilities of once-future things. *Organizational Behavior and Human Performance, 13*, 1-16.

Forster, J., & Strack, F. (1996). Influence of overt head movements on memory for valenced words: A case of conceptual-motor compatibility. *Journal of Personality and Social Psychology, 71*, 421-430.

Forster, J., & Strack, F. (1997). Motor actions in retrieval of valenced information: A motor congruence effect. *Perceptual and Motor Skills, 85*, 1419-1427.

Friedman, R. S., & Foerster, J. (2000). The effects of approach and avoidance motor actions on the elements of creative insight. *Journal of Personality and Social Psychology, 79*, 477-492.

Friedman, R. S., & Foerster, J. (2002). The influence of approach and avoidance motor actions on creative cognition. *Journal of Experimental Social Psychology, 38*, 41-55.

Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189-211). New York: Guilford.

Gilbert, D. T. (2002). Inferential correction. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 167-184). Cambridge, UK: Cambridge University Press.

Gilovich, T., & Medvec, V. H. (1995). The experience of regret: What, when, and why. *Psychological Review, 102*, 379-395.

Gilovich, T., Medvec, V. H., & Savitsky, K. (2000). The spotlight effect in social judgment: An egocentric bias in estimates of the salience of one's own actions and appearance. *Journal of Personality and Social Psychology, 78*, 211-222.

Gilovich, T., & Savitsky, K. (1999). The spotlight effect and the illusion of transparency: Egocentric assessments of how we're seen by others. *Current Directions in Psychological Science, 8*, 165-168.

Gilovich, T., Savitsky, K., & Medvec, V. H. (1998). The illusion of transparency: Biased assessments of others' ability to read our emotional states. *Journal of Personality and Social Psychology, 75*, 332-346.

Hawkins, S. A., & Hastie, R. (1991). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin, 107*, 311-327.

Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin, 21*, 1161-1167.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201-208). Cambridge, UK: Cambridge University Press.

Keysar, B., & Barr, D. J. (2002). Self-anchoring in conversation: Why language users don't do what they "should." In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 150-166). Cambridge, UK: Cambridge University Press.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology, 89*, 46-55.

Logan, G. D. (1997). The automaticity of academic life: Unconscious applications of an implicit theory. In R. S. Wyer (Ed.), *The automaticity of everyday life: Advances in social cognition* (Vol. 10, pp. 157-179). Mahwah, NJ: Lawrence Erlbaum.

Martin, L. L., Ward, D. W., Achee, J. W., & Wyer, R. S. (1993). Mood as input: People have to interpret the motivational implications of their moods. *Journal of Personality and Social Psychology, 64*, 317-326.

Miller, D. T., & Taylor, B. R. (1995). Counterfactual thought, regret, and superstition: How to avoid kicking yourself. In N. J. Roese & J. M. Olson (Eds.), *What might have been: The social psychology of counterfactual thinking* (pp. 305-331). Hillsdale, NJ: Lawrence Erlbaum.

Mussweiler, T., & Strack, F. (1999a). Comparing is believing: A selective accessibility model of judgmental anchoring. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 10). Chichester, UK: Wiley.

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

Mussweiler, T., & Strack, F. (2000). The use of category and exemplar knowledge in the solution of anchoring tasks. *Journal of Personality and Social Psychology, 78*, 1038-1052.

Mussweiler, T., Strack, F., & Pfeiffer, T. (2000). Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility. *Personality and Social Psychology Bulletin, 26*, 1142-1150.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259.

Plous, S. (1989). Thinking the unthinkable: The effect of anchoring on likelihood estimates of nuclear war. *Journal of Applied Social Psychology, 19*, 67-91.

Quattrone, G. A. (1982). Overattribution and unit formation: When behavior engulfs the person. *Journal of Personality and Social Psychology, 42*, 593-607.

Quattrone, G. A., Lawrence, C. P., Finkel, S. E., & Andrus, D. C. (1981). *Explorations in anchoring: The effects of prior range, anchor extremity, and suggestive hints.* Unpublished manuscript, Stanford University.

Rottenstreich, Y., & Tversky, A. (1997). Unpacking, repacking, and anchoring: Advances in support theory. *Psychological Review, 104*, 406-415.

Schkade, D. A., & Johnson, E. J. (1989). Cognitive processes in preference reversals. *Organizational Behavior and Human Decision Processes, 44*, 203-231.

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

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review, 93*, 239-257.

Trope, Y., & Gaunt, R. (2000). Processing alternative explanations of behavior: Correction or integration? *Journal of Personality and Social Psychology, 79*, 344-354.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1130.

Wegener, D. T., & Petty, R. E. (1995). Flexible correction processes in social judgment: The role of naïve theories in corrections for perceived bias. *Journal of Personality and Social Psychology, 68*, 36-51.

Wegener, D. T., Petty, R. E., Detweiler-Bedell, B. T., & Jarvis, W. B. G. (2001). Implications of attitude change theories for numerical anchoring: Anchor plausibility and the limits of anchor effectiveness. *Journal of Experimental Social Psychology, 37*, 62-69.

Wells, G. L., & Petty, R. E. (1980). The effects of overt head movements on persuasion: Compatibility and incompatibility of responses. *Basic and Applied Social Psychology, 1*, 219-230.

Wilson, T. D., Houston, C., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology: General, 4*, 387-402.

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 & Human Decision Processes, 82*, 314-333.

Wright, W. F., & Anderson, U. (1989). Effects of situation familiarity and incentives on use of the anchoring and adjustment heuristic for probability assessment. *Organizational Behavior and Human Decision Processes, 44*, 68-82.

Received January 17, 2003
Revision accepted June 14, 2003