Journal of Experimental Social Psychology **35,** 136–164 (1999)
Article ID jesp.1998.1364, available online at http://www.idealibrary.com on IDEAL®

Hypothesis-Consistent Testing and Semantic Priming in the Anchoring Paradigm: A Selective Accessibility Model

Thomas Mussweiler

Northwestern University

and

Fritz Strack

Universität Würzburg, Würzburg, Germany

Received June 11, 1998; Revised July 3, 1998; Accepted July 15, 1998

Results of four studies support the notion that anchoring effects are mediated by mechanisms of hypothesis-consistent testing and semantic priming. According to the suggested Selective Accessibility Model, judges use a hypothesis-consistent test strategy to solve a comparative anchoring task. Applying this strategy selectively increases the accessibility of anchor-consistent knowledge which is then used to generate the subsequent absolute judgment. Studies 1 and 2 demonstrate that absolute estimates depend on the hypothesis implied in the comparative task, suggesting that a hypothesis-testing strategy is used to solve this task. Study 3 shows that limiting the amount of knowledge generated for the comparative task retards absolute judgments. This suggests that knowledge rendered easily accessible in the comparative judgment is used for the subsequent absolute judgment. Finally, Study 4 suggests that self-generation of knowledge contributes to the robustness of the effect, thus resolving the seeming inconsistency that anchoring effects are at the same time remarkably robust and mediated by typically fragile semantic priming mechanisms.

Key Words: anchoring heuristic; selective accessibility; hypothesis-consistent testing; semantic priming.

The present research was supported by a grant from the Deutsche Forschungsgemeinschaft. We thank Russell Fazio, John Skowronski, Timothy Wilson, Bob Wyer, and an anymous reviewer for helpful and extensive comments on an earlier draft of this manuscript. Thanks also go to Jens Förster, Roland Neumann, Sabine Stepper and Lioba Werth for discussions of the issue, to Markus Denzler, Katrin Meier, Amina Özelsel, Markus Schuhmacher, Bärbel Schöppner, Anke Siebers, and Marion Weiss for their assistance in data collection and to Jens Unger for the computer programming. Finally, we want to thank Gerd Bohner for giving methodological advice.

Address correspondence and reprint requests to Thomas Mussweiler, Department of Psychology, Northwestern University, 2029 Sheridan Road, Evanston, IL 60208-2710. E-mail: t-mussweiler@uwu.edu.

0022-1031/99 $30.00

Some 25 years ago, Tversky and Kahneman (1973, 1974) laid the foundations for what became one of the most influential research programs in psychology: the heuristics and biases approach. Despite its significant contribution to a broad array of scientific fields, this approach has been fiercely criticized (e.g. Cohen, 1981; Einhorn & Hogarth, 1981; Gigerenzer, 1996; Lopes, 1991). One of the central criticisms brought forward is that research has focused almost exclusively on the judgmental effects of heuristics and biases while neglecting their underlying processes. This critique, however, does not equally apply to the three classic heuristics of *availability, representativeness* and *anchoring.* Some insights into the processes that underlie the availability (cf. Schwarz, Bless, Strack, Klumpp, Rittenauer-Schatka, & Simons, 1991) and the representativeness heuristic (cf. Tversky, 1977; Tversky & Gati, 1978) have been gained, whereas the processes that underlie the anchoring heuristic remain unclear (Strack & Mussweiler, 1997).

Anchoring is apparent in the assimilation of a numeric estimate to a previously provided standard. In what is probably the best known demonstration of this effect (Tversky & Kahneman, 1974), estimates for the percentage of African nations in the UN were assimilated toward a given standard, so that high standards yielded higher estimates than low standards. Anchoring effects like these have proved to be a truly ubiquitous and robust phenomenon. In particular, they have been observed in a broad array of different judgmental domains, such as general knowledge questions (Jacowitz & Kahneman, 1995; Strack & Mussweiler, 1997; Tversky & Kahneman, 1974; Wilson, Houston, Etling, & Brekke, 1996), real estate evaluation (Northcraft & Neale, 1987), estimates of self efficacy (Cervone & Peake, 1986; Switzer & Sniezek, 1991), probability assessments (Holtgraves & Skeel, 1992; Joyce & Biddle, 1981; Plous, 1989; Tversky & Kahneman, 1974; Wright & Anderson, 1989), and evaluations of lotteries and gambles (Carlson, 1990; Chapman & Johnson, 1994; Johnson & Schkade, 1989; Schkade & Johnson, 1989). In addition, anchoring remains uninfluenced by the extremity of the anchor (Chapman & Johnson, 1994; Quattrone, Lawrence, Warren, Souza-Silva, Finkel, & Andrus, 1984; Strack & Mussweiler, 1997), increased motivation, correctional instructions (Wilson et al., 1996) and participants' expertise[1] (Joyce & Biddle, 1981; Northcraft & Neale, 1987; Wright & Anderson, 1989).

Moreover, anchoring constitutes a basic explanatory concept that has been applied to conceptualize a variety of judgmental phenomena, such as the correspondence bias (Leyens, Yzerbyt, & Corneille, 1996; Quattrone, 1982), the hindsight bias (Fischhoff, 1975; Pohl, 1996), preference reversal effects (Lichtenstein & Slovic, 1971; Schkade & Johnson, 1989) and probabilistic inferences (Carlson, 1990; Einhorn & Hogarth, 1981; Svenson, 1985; Tversky & Kahneman, 1974).

In marked contrast to this exceptional empirical and theoretical significance, little is known about the processes that lead to anchoring. To overcome this

[1] See Mussweiler and Strack (1998) for a discussion of the disparate effects of judges' knowledgeability and expertise on the anchoring effect.

deficiency, we proposed that anchoring effects are mediated by mechanisms of selective accessibility (Strack & Mussweiler, 1997). The present paper further specifies this possibility and provides more unequivocal support for our conceptualization. Specifically, it demonstrates that anchoring effects may be based on a mechanism that combines two fundamental notions of social cognition research: *hypothesis-consistent testing* and *semantic priming.*

THE SELECTIVE ACCESSIBILITY MODEL

Recently, we (Strack & Mussweiler, 1997; see also Mussweiler & Strack, in press) have suggested a *Selective Accessibility Model* to account for anchoring phenomena. The model recognizes that in the standard anchoring paradigm, participants have to perform two consecutive tasks: a comparative judgment and an absolute judgment. First, in the *comparative task,* participants are requested to compare the target object with a given standard, the anchor. Thus, in the example mentioned before, Tversky and Kahneman (1974) first asked their participants whether the percentage of African nations in the UN was higher or lower than an arbitrary number (the anchor) that had ostensibly been determined by spinning a wheel of fortune (e.g., 65% or 10%). In the second *absolute task,* participants were then asked to give their best estimate of this percentage. Absolute judgments were assimilated to the anchor that was salient in the comparative judgment task, so that the mean estimate of participants who had received the high anchor was 45%, compared to 25% for participants who had received the low anchor.

According to the Selective Accessibility Model, participants solve the comparative task by selectively generating semantic knowledge that is consistent with the notion that the target's value is equal to the anchor (the *Selectivity Hypothesis*). Generating such knowledge increases its subsequent accessibility, so that it is used to form the final absolute judgment (the *Accessibility Hypothesis*).

Hypothesis-Consistent Testing: The Selectivity Hypothesis

Research on hypothesis testing (Snyder & Swann, 1978; Wason, 1960; Wason & Johnson-Laird, 1972) has demonstrated that judges often test a given hypothesis by focusing primarily on consistent evidence. This preference for hypothesis-consistent evidence is often adaptive (Klayman & Ha, 1987; Trope & Liberman, 1996). That is, adopting a *positive test strategy* (Klayman & Ha, 1987) (i.e., examining instances in which the target characteristic is present) is often the most critical test of the hypothesis under consideration.

In linking this body of research to the anchoring paradigm, we (Strack & Mussweiler, 1997) suggested that participants solve the comparative task by using such a hypothesis-testing strategy. More specifically, participants who are asked whether the target's value along the judgmental dimension is greater or smaller than the anchor value may generate the answer by testing the hypothesis that the target's value is equal to the anchor. For example, participants who are asked whether the annual mean temperature in Germany is higher or lower than 20°C

may generate the answer by testing the possibility that the mean temperature is 20°C.[2]

In line with the notion that hypothesis-consistent testing is a reasonable strategy in many situations, we further assume that this focal hypothesis is tested by adopting such a strategy (Klayman & Ha, 1987). Hence, participants may test the focal hypothesis by examining the possibility that the target object's value along the judgmental dimension is equal to the anchor value. In order to do so, participants may try to generate a mental model (Johnson-Laird, 1983) of the target with the extension of the anchor by selectively retrieving knowledge from memory that is consistent with this notion. That is, they may initiate a selective search for hypothesis-consistent evidence. In our example, they may try to retrieve knowledge that implies that the annual mean temperature in Germany is indeed 20°C. Thus, they may recall that in summer the temperatures even exceed 20°C, that already in spring peak temperatures sometimes are about 20°C, that people wear shorts and short sleeves a lot, etc.

Semantic Priming: The Accessibility Hypothesis

How does considering the focal hypothesis of the comparative question produce the anchoring effect on the subsequent absolute judgment? The Selective Accessibility Model assumes that the process mediating this effect is akin to semantic priming. Research on semantic priming has repeatedly demonstrated the effects of activating knowledge on a subsequent judgment (for recent reviews see Higgins, 1989, 1996, 1997; Sedikides & Skowronski, 1991; Wyer & Srull, 1989). For example, in the context of trait ascription (e.g., Higgins, Rholes, & Jones, 1977; Srull & Wyer, 1979), participants have been found to use trait concepts which were primed in a preceding unrelated task to characterize an ambiguously described target person (''Donald'').

[2] Although other mechanisms may well be involved in the generation of the comparative judgment, the suggested hypothesis-testing mechanism is consistent with both the literature on social hypothesis testing (for a review, see Trope & Liberman, 1996) and the evidence reported in the present manuscript. It has been demonstrated that in most cases judges seek evidence that is diagnostic for the judgment at hand (e.g., Trope & Bassok, 1982). For the comparative judgment, evidence is diagnostic if it allows to decide whether the target's extension is higher or lower than the anchor value. With limited knowledge about the judgmental target, participants searching for such diagnostic evidence are likely to entertain the possibility that the target's value is equal to the anchor value. Thus, they may transform the comparative judgment into the test of a single focal hypothesis. As Trope and Liberman (1996) point out, such single hypothesis-testing is diagnostic if no specific alternative hypothesis is provided. Note that in the current context the focal hypothesis of the comparative question (e.g., ''The annual mean temperature in Germany is 20°C'') is more specific than its alternative (''The annual mean temperature in Germany is *not* 20°C''). Thus, participants cannot test the alternative hypothesis per se. They can only test *one* of all possible alternatives (e.g., ''The annual mean temperature in Germany is *10°C*''). However, whereas for direct tests of the focal hypothesis confirmation and disconfirmation are diagnostic for the judgment at hand, for the alternative hypothesis only confirmation is diagnostic. Consequently, testing the focal hypothesis of the comparative question appears to be an effective way to solve the comparative task.

Generating knowledge in order to solve the comparative task may influence the subsequent absolute judgment in much the same way. In particular, generating knowledge increases its subsequent accessibility, so that it is more likely to be used for the absolute judgment. However, in adopting a positive test strategy, participants do not generate a representative set of knowledge for the target. Rather, they recall knowledge selectively. Thus, solving the comparative task selectively increases the accessibility of anchor-consistent knowledge. When solving the absolute task, participants resort to this easily accessible knowledge and therefore base their absolute estimate primarily on anchor-consistent evidence. This manifests itself in an assimilation of the absolute estimate to the anchor value.

In sum, the Selective Accessibility Model is built on two fundamental hypotheses. First, participants are assumed to answer the comparative question of an anchoring task by testing its focal hypothesis. In order to do so, they apply a hypothesis-consistent test strategy (Selectivity Hypothesis) which leads to a selective increase in the accessibility of anchor-consistent evidence. Second, to generate the subsequent absolute judgment participants are assumed to resort to this easily accessible evidence (Accessibility Hypothesis), which leads to the assimilation of the absolute response to the anchor provided in the comparative question.

Notably, this selective accessibility mechanism is also consistent with explanations of the hindsight phenomenon, where a recalled judgment is assimilated to a provided outcome (for a review, see Hawkins & Hastie, 1990). For example, Hasher, Attig, and Alba (1981) assume that the hindsight bias is mediated by the increased accessibility of outcome-congruent information (see also Pohl, 1996). Chapman and Johnson (1994) have proposed that a similar mechanism may play a role in anchoring. Specifically, they suggest that ''. . . the presence of an anchor increases the availability of features that the anchor and target hold in common . . .'' (p. 239).

Evidence for the Selective Accessibility Model

A series of studies (Mussweiler, Fō rster, & Strack, 1997; Mussweiler & Strack, 1998; Strack & Mussweiler, 1997; for a review, see Mussweiler & Strack, in press) provided support for these assumptions. The most compelling evidence stems from an experiment which combined the anchoring paradigm with a lexical decision task (Mussweiler & Strack, 1998). Specifically, we found that on a lexical decision task that followed the comparative question, participants were faster in identifying anchor-consistent words than anchor-inconsistent words. For example, participants who had just decided whether the average price for a German car is higher or lower than 40,000 German Marks (i.e. the high anchor) were faster in identifying words associated with expensive cars (e.g., ''Mercedes,'' ''BMW'') than words associated with cheap cars (e.g., ''Golf,'' ''Volkswagen''). In contrast, for participants who indicated whether the average price is higher or lower than 20,000 German Marks (i.e., the low anchor), the

reverse was true. This pattern of response latencies for lexical decisions suggests that solving the comparative anchoring task selectively increases the accessibility of anchor-consistent knowledge.

Additional evidence further suggests that judges base their absolute estimate primarily on evidence that has been rendered easily accessible during the comparative task. Suppose, this is indeed the case. Then, the time that is needed to generate the absolute estimate should depend on the accessibility of judgment-relevant knowledge. Specifically, the more easily accessible the relevant evidence is, the faster the absolute question should be answered.

In order to test this assumption, we manipulated the amount of knowledge generated for the comparative test by varying the plausibility of the anchors (Strack & Mussweiler, 1997, Study 3). Note that plausible anchors require the generation of more knowledge about the target than implausible ones. For example, deciding whether the German river Elbe is longer or shorter than 890 kilometers is more difficult and thus requires more knowledge than deciding whether it is longer or shorter than 45,000 kilometers. Accordingly, the accessibility of judgment-relevant knowledge should be greater so that absolute estimates can be generated faster after comparing the Elbe to 890 kilometers than after comparing it to 45,000 kilometers. Consistent with this assumption, response latencies for the absolute judgment (e.g., ''How long is the river Elbe?'') were shorter when the anchor mentioned in the comparative question was plausible than when it was implausible.

Taken together, these results suggest that solving a comparative anchoring task selectively increases the accessibility of anchor-consistent evidence which is then used to generate the absolute estimate. Although our initial findings are thus consistent with the Selective Accessibility Model, however, they leave a number of critical questions unanswered and do not provide unequivocal support for our conceptualization. For one, they do not specify the mechanisms that lead to the selective increase in the accessibility of anchor-consistent knowledge. The Selective Accessibility Model explicitly assumes that testing the hypothesis that the target's value is equal to the anchor value is responsible for this increase. Studies 1 and 2 were designed to test this possibility. Moreover, the evidence suggesting that absolute estimates are based on knowledge that has been rendered easily accessible during the comparative task is indirect, because accessibility was manipulated via the plausibility of the anchor values. Study 3 attempts to provide more direct support for this assumption. Finally, the notion that anchoring effects are mediated by a mechanism that is akin to semantic priming seems problematic, because priming effects and anchoring effects differ with regard to one important characteristic, namely their robustness. As mentioned before, anchoring effects have proved to be exceptionally robust. In contrast, priming effects are rather fragile (for a discussion, see Strack, 1992). If both phenomena were mediated by the same mechanism, this disparity needs to be explained. Study 4 explicitly tests such an explanation.

STUDY 1

Suppose participants do indeed solve a comparative anchoring task by testing the hypothesis implied in it. Then, different hypotheses should yield different tests and ultimately lead to different estimates. Study 1 investigated this possibility.

In particular, we manipulated the focal hypothesis by subtly changing the wording of the question. Participants were either asked to indicate whether the target object is *larger* than the anchor value or they were asked whether it is *smaller* than this value. For example, participants were either asked whether the river Elbe is *longer* than 890 kilometers, or they were asked whether the River Elbe is *shorter* than this value. If our reasoning is correct, participants should test for different possibilities in both conditions. Presupposing that a strategy of hypothesis-consistent testing is used, these different hypotheses should initiate a search for evidence in opposite directions. While participants should selectively search for evidence implying that the river Elbe is longer than 890 kilometers in the first case, they should search for evidence implying that the river Elbe is shorter than 890 kilometers in the latter case. As a result, absolute estimates should be higher in the ''longer'' condition than in the ''shorter'' condition.

Moreover, the notion that participants apply a hypothesis-consistent test strategy to solve the comparative anchoring task implies that the assimilative influence of this comparison should already be apparent in the comparative judgment itself. If judges solve the comparative task by selectively generating anchor-consistent evidence and subsequently base their comparative judgment on this evidence, then their comparative judgment should be consistent with the generated evidence. As a consequence, judges should affirm the possibility suggested in the comparative question more often than a control group that did not make the comparison (i.e., received absolute questions only). For example, judges asked whether the river Elbe is longer than 890 kilometers should affirm this possibility more often than judges asked: ''How long is the Elbe?''

Jacowitz and Kahneman (1995) provided some descriptive evidence that supports this possibility. Specifically, they found that although only 15% of participants who did not make a comparison with a specific high anchor value (i.e., received the absolute question only) stated that the actual value is higher than the anchor, 27% of participants who made this comparison did so. Thus, comparing the target to a high anchor value increased the probability that participants saw the target as even larger. This assimilative influence on comparative judgments, however, was only obtained for high, not for low anchors, which is difficult to explain on a priori grounds. In order to demonstrate the generality as well as the statistical reliability of this effect, Study 1 also explores the effects that comparing the target to the anchor has on the comparative judgment.

Method

Participants. We recruited 39 male and female non-psychology students of the University of Würzburg as participants. They were asked to take part in a pretest for the construction of a questionnaire assessing general knowledge and were offered a chocolate bar as compensation.

TABLE 1
OBJECTS AND ANCHORS USED IN STUDIES 1 AND 2

Question	Actual value	High anchor	Low anchor
Antarctic: mean temperature in winter (°C)	−68	−17	−43
Einstein: year of first visit to USA	1921	1939	1905
Da Vinci: year of birth	1452	1698	1391
Gandhi: age	78	79	64
Ulm: altitude (m)	478	320	150
Aristotle: year of birth	−322	−220	−490
Whale: length (m)	33	49	21
Elbe: length (km)	1165	890	550

Materials. The questionnaire consisted of 8 pairs of comparative and absolute questions. The questions used were similar to those of Strack and Mussweiler (1997, Study 3). The anchors were either one standard deviation higher or one standard deviation lower than the mean estimates of a calibration group ($N = 151$). Half of the questions included a high anchor, the other half included a low anchor.

In line with the above reasoning, the focal hypothesis of the comparative question was varied: Half of the comparative questions asked whether the target is *larger* than the anchor value, while the other half asked whether it is *smaller* than this value. The response alternatives provided in either case were ''yes'' and ''no.'' For example, participants were either asked whether the river Elbe is *longer* than 890 kilometers or they were asked whether it is *shorter* than 890 kilometers. Then, they were given the absolute question (e.g., ''How long is the Elbe?''). Thus, the four experimental conditions resulted from a combination of high versus low anchors and ''larger'' versus ''smaller'' question wording. The targets and anchors are listed in Table 1.

A Latin-Square design was applied to control for content and order effects. Four different versions of the questionnaire were constructed. In all of these, questions were presented in the same order depicted in Table 1. However, in each version, the different conditions were assigned to different questions, so that over all versions each of the conditions was realized with each of the eight critical question pairs. In addition, we counterbalanced the order of the conditions.

Procedure. Participants were recruited in the university cafeteria and were then led to a separate room in which they completed the questionnaire in groups of up to 15. Upon arrival, they were given the questionnaire and were told to read instructions carefully. They were informed that they were taking part in a pretest for the construction of a general-knowledge questionnaire. The purpose of the pretest was ostensibly to find the best wording for general-knowledge questions. To reduce the prescribed informativeness of the anchors and thus discourage conversational inferences (Grice, 1975), participants were told that the values were randomly selected.[3] It was pointed out that this random selection of the anchors was necessary to minimize their impact on the answers and to identify the impact of different question formats. Finally, participants were instructed to answer the questions as accurately as possible.

[3] It has been suggested (e.g., Jacowitz & Kahneman, 1995) that applying implicit rules of natural conversations (Grice, 1975) to standardized situations (e.g., Clark & Schober, 1992; Schwarz, 1994; Strack & Martin, 1987) allows participants to use the anchor value in order to infer the actual range of possible answers. Participants who expect the experimenter to be maximally informative (see Grice's, 1975, *maxim of quantity*) in asking his or her questions, may assume that the anchor value is close to the actual value and consequently position their estimate in its vicinity. This explanation, however, presupposes that the anchor value is deliberately selected by the experimenter. Thus, conversational inferences cannot explain the effects of randomly selected anchor values.

Results

Comparative judgments. To explore whether the comparative judgments were influenced by the anchor values, they were compared to the estimates of the calibration group. To do so, we calculated deviation scores for each comparative answer. ''Correct'' responses to the comparative question (''No'' for ''Is the target value larger than the high anchor'' or ''Is the target value smaller than the low anchor'') were coded 0, ''incorrect'' responses (e.g., ''Yes'' for ''Is the target value larger than the high anchor'') were coded 1. As described above, high and low anchors deviated from the mean of the calibration group by one standard deviation, so that for each of the 8 questions about 16% of the calibration participants stated that the target's value is smaller than the low anchor and another 16% stated that it is larger than the high anchor. Thus, with respect to each anchor value, 16% of the calibration subjects gave ''incorrect'' answers. If comparative answers remained uninfluenced by the anchor value, the same proportion of ''incorrect'' answers should be obtained, so that the expected value for each of the comparative questions was .16.

As a measure for the deviation of comparative judgments from their expected value, we subtracted the expected value (i.e., .16) from each of the given comparative answers. Overall, the mean deviation score was .21, indicating that experimental participants gave 21% more ''incorrect'' answers than calibration participants, $F(1, 38) = 28.02$, $p < .001$.[4] The magnitude of this deviation did not depend on the anchor ($M = .25$ vs $M = .16$, for high an low anchors respectively), $F(1, 38) = 1.26$, $p > .25$, for the main effect of Anchor. Moreover, it was not influenced by the wording of the comparative question ($M = .19$ vs $M = .21$, for ''larger'' and ''smaller'' respectively), $F(1, 38) < 1$, for the remaining effects.

Absolute estimates. Two participants had to be excluded from the analysis, because they did not answer all 8 absolute questions of the questionnaire. Thus the analysis of the absolute estimates is based on the responses of the remaining 37 participants. To pool answers across different content domains, absolute estimates were transformed into *z*-scores. Thus, the resulting scores reflect participants' average deviation from the question mean in units of the pertinent standard deviation.

High anchors led to higher absolute estimates ($M = .24$) than low anchors ($M = -.37$), $F(1, 36) = 39.21$, $p < .001$. Independently of this effect, however, absolute estimates were influenced by whether the comparative question asked respondents to determine whether the target was larger than the anchor or smaller than the anchor. Specifically, higher estimates were made for the ''larger'' question ($M = .04$) than for the ''smaller'' question ($M = -.17$), $F(1, 36) = 5.72$, $p < .02$. This difference did not depend on whether the anchor was high (.32 vs .16) or low (−.25 vs −.49), $F(1, 36) < 1$, for the interaction.

[4] The described deviation scores were used as dependent variables in a repeated measures ANOVA. As Rosenthal and Rosnow (1985) point out, this is an appropriate method to analyze proportions ranging from .15 to .85.

Discussion

The implications of these findings are two-fold. First, the analysis of the comparative judgments indicates that participants in all experimental conditions gave significantly more ''incorrect'' answers than would be expected on the basis of our calibration data. That is, experimental subjects stated that the target is *larger* than the *high anchor* or *smaller* than the *low anchor* more often than calibration subjects. In contrast to Jacowitz and Kahneman (1995), this effect occurred for high as well as for low anchors. Thus, the assimilative influence of the anchor value is not restricted to the absolute judgment. Rather, comparing the target to the anchor value influences comparative judgments already, so that they are more consistent with the possibility that was suggested in the comparative question. This finding provides additional support for the Selectivity Hypothesis. It indicates that in order to compare the target to the anchor value, judges engage in hypothesis-consistent testing which fosters their belief in the possibility that is suggested by the comparison.

More important, the absolute estimates depended on the direction of search that was stimulated by attempts to test the hypothesis implied by the comparative question. The fact that absolute estimates are sensitive to such subtle changes in the focus of the implied hypothesis suggests that anchoring is indeed mediated by a hypothesis-testing process.

However, our conceptualization of anchoring is more specific with regard to the supposed hypothesis-testing process. In the standard anchoring paradigm (e.g., Tversky & Kahneman, 1974), comparative questions asked participants to indicate whether the target is *larger or smaller* than the anchor value. In this case, we assume that participants test the hypothesis that the target is *equal* to the anchor. If this is the case, similar tests leading to similar absolute estimates should be initiated for the standard comparative question and a comparative question explicitly asking whether the target's extension is about equal to the anchor value. This reasoning was tested in Study 2.

STUDY 2

Method

Participants. We recruited 40 male and female non-psychology students of the University of Würzburg as participants.

Materials. Except for the wording of the comparative question, the materials were identical to those used in Study 2. In line with the above reasoning, half of the comparative questions asked whether the target is larger or smaller than the anchor value, while the other half asked whether the extension of the target is about equal to the anchor. For example, participants were either asked whether the river Elbe is longer or shorter than 890 kilometers, or they were asked, whether the river Elbe is about 890 kilometers long.

Procedure. The procedure was identical to Study 1.

Power analysis. Because our prediction for the current study consists of a null hypothesis, we conducted a power analysis to test for the reliability of the expected results. For $\alpha = .05$ and $N = 40$, the power to detect a medium-sized effect ($d = .5$) was $1 - \beta > .9$ (cf. Cohen, 1977).

Results

One participant was excluded from the analysis of the comparative as well as the absolute judgments, because he did not answer all comparative questions.

Comparative judgments. To explore whether the comparative judgments were influenced by the anchor values, we calculated the same deviation scores as in Study 1. Overall, the mean score was $M = .39$, indicating that experimental participants gave 39% more ''incorrect'' answers to the comparative question than calibration participants, $F(1, 38) = 89.95$, $p < .001$. Again, this proportion did not differ for the high ($M = .44$) and the low anchor ($M = .34$), $F(1, 38) = 2.56$, $p > .12$. Moreover, it remained uninfluenced by the wording of the comparative question ($M = .40$ vs $M = .38$ for ''larger or smaller'' and $M = .48$ vs $M = .30$ for ''about equal''), $F(1, 38) < 1$, for all remaining effects.

Absolute estimates. Three participants had to be excluded from the analysis, because they did not answer all of the absolute questions. Consequently, the analysis of the absolute estimates is based on the responses of the remaining 36 participants.

Again, using *z*-transformed absolute estimates as dependent variable, high anchors led to higher estimates ($M = .28$) than low anchors ($M = -.19$), $F(1, 35) = 26.07$, $p < .001$. However, absolute estimates did not depend on the wording of the comparative question. Specifically, similar estimates were given when the comparative question asked whether the target was larger or smaller than the anchor ($M = .04$) and when it asked whether the target was equal to the anchor ($M = .06$), $F(1, 35) < 1$. Moreover, there was no interaction of Anchor and Question Wording, $F(1, 35) < 1$.

To examine whether the dispersion of the absolute estimates differed for the two question formats, we calculated individual deviation scores for each absolute estimate by subtracting the mean for the respective experimental condition. Deviation scores for the ''about equal'' question ($M = .43$) and the ''larger or smaller'' question ($M = .40$) did not differ, $F(1, 35) < 1$, indicating similar dispersions of absolute values in both question formats.

Discussion

The fact that absolute estimates were similar for comparative questions that explicitly asked whether the target is about the size of the anchor value and for the standard comparative question suggests that similar tests were performed in both conditions. Thus, participants may solve the standard comparative task by testing the possibility that the target's value is equal to the anchor value. However, it is important to note that, in principle, the obtained similarity of absolute estimates may also result because participants either test whether the target is larger than the anchor value, or whether it is smaller than this value. Aggregating data over participants testing either of these hypotheses may then offset the effects of the knowledge activated in both cases. This, however, should lead to a wider dispersion of absolute estimates for the standard comparative question. Thus, the fact that dispersions were almost identical argues against this possibility and

further supports the assumption that participants solve the standard comparative question by testing the possibility that the target's value is equal to the anchor value.

Taken together, the results of Studies 1 and 2 help specify the mechanisms responsible for the selective increase in the accessibility of anchor-consistent knowledge that appears to result from solving a comparative anchoring task. To compare the target to the anchor value, judges appear to test the hypothesis that is implied in the comparative question. Thus, in the standard anchoring paradigm, they may test for the possibility that the target's value is equal to the anchor value. To do so, they appear to engage in hypothesis-consistent testing, which fosters their belief in the tested possibility. In the present data this is apparent in the comparative as well as the absolute judgments.

How does the selective increase in the accessibility of anchor-consistent knowledge that results from the described hypothesis test influence the subsequent absolute judgment? As lined out before, the Selective Accessibility Model assumes that the absolute judgment is influenced, because participants resort primarily to easily accessible knowledge to generate this judgment (see the Accessibility Hypothesis). Support for this assumption stems from a study investigating whether the time participants need to solve the absolute judgment depends on the plausibility of the comparison standard (Strack & Mussweiler, 1997, Study 3). Comparing the target to a plausible standard requires more time presumably because more knowledge has to be generated to make the judgment. As a consequence, more knowledge that is relevant for the absolute task is rendered easily accessible, so that absolute estimates should be facilitated. In line with this reasoning, we found that response latencies for the absolute anchoring question depended on the plausibility of the anchor values. If judges needed longer to solve the comparative task because it included a plausible rather than an implausible anchor value, they generated the subsequent absolute estimate faster. This suggests that absolute estimates are based on evidence that was rendered easily accessible in the preceding comparative task as is proposed by the Accessibility Hypothesis.

These data, however, only provide fairly indirect support for the Accessibility Hypothesis because the accessibility of judgment-relevant knowledge was manipulated via the plausibility of the anchor values. Alternatively, accessibility may be manipulated in a more direct fashion, by varying the time participants are given to solve the comparative task. Time pressure heightens need for closure (Kruglanski & Webster, 1996) which in turn decreases the amount of knowledge that people consider before forming a judgment (cf. Mayseless & Kruglanski, 1987; see also Strack, Erber, & Wicklund, 1982).

Similarly, participants who feel to be under time pressure while working on the comparative task are likely to generate less knowledge. If this is the case, time pressure should decrease the amount of evidence that is activated in performing the task and thus the amount that is easily accessible at the time the subsequent absolute judgment is made. Therefore, putting participants under time pressure

TABLE 2
OBJECTS AND ANCHORS USED IN STUDY 3

Question	Actual value	Plausible anchors		Implausible anchors	
		High	Low	High	Low
Antarctic: mean temperature in winter (°C)	−68	−17	−43	+45	−210
Da Vinci: year of birth	1452	1698	1391	1952	300 B.C.
Einstein: year of first visit to USA	1921	1939	1905	1992	1215
Elbe: length (km)	1165	890	550	45000	25

during the comparative judgment task should increase the time required to generate the absolute judgment. These implications were tested in Study 3.

STUDY 3

Method

Participants. 42 students of the University of Würzburg were recruited as participants. As in the preceding studies, they were asked to take part in a pretest for the construction of a general knowledge questionnaire and were offered a chocolate bar as compensation.

Materials. The questions used were similar to those of Study 1. In particular, the questions and anchors were chosen from those used by Strack and Mussweiler (1997, Study 3). Accordingly, the anchors differed in their direction and their plausibility. Plausible anchors deviated by about one standard deviation from the mean of a calibration group ($N = 151$); implausible anchors deviated from this mean by more than 10 standard deviations unless such an extreme deviation yielded logical inconsistencies. In addition, the plausibility was rated by 40 different participants (cf. Strack & Mussweiler, 1997). Thus, four different anchor types resulted from the orthogonal combination of plausibility (plausible vs implausible) and anchor (high vs low). The questions and anchors used are listed in Table 2. Again, content and order effects were controlled, using a Latin-Square design. Thus, each type of anchor was used in all question pairs, whereas question order was kept constant.

Procedure. Participants took part in the experiment in groups of up to four. They were recruited in the university cafeteria, escorted to the computers and told to read instructions carefully. General instructions were identical to those used in the preceding studies. In addition, participants in the time-pressure condition were informed that they would have five seconds to answer the comparative question and that the computer would proceed automatically after these five seconds had elapsed. In our earlier study (Strack & Mussweiler, 1997, Study 3), participants needed an average of about six seconds to answer the comparative question. Thus, the time pressure exerted was fairly mild, leaving enough time to process the question. Instructions for the participants in the no-time-pressure condition did not include the paragraph pertaining to the time-pressure manipulation.

After participants had read the instructions, the experimenter demonstrated how to report answers using the keyboard. Participants were told to answer the comparative question by pressing either the *q*-key, which was marked with a red sticker or the *p*-key, which was marked with a green sticker. For each comparative question the keys corresponding to the two possible answers (e.g., longer or shorter) were depicted on the bottom of the computer screen. In order to reduce variance in response latencies, participants were told to position their forefingers on the two keys before the question appeared on the screen. Participants were told that to answer the absolute questions they should use the number pad on the keyboard. They were warned that comparative and absolute questions would alternate. Finally, participants were instructed to answer the questions as accurately and as fast as possible.

Subsequently, participants were presented 17 pairs of comparative and absolute questions. The first 13 pairs served as practice trials. Pairs 14 through 17 were the critical trials and were thus included in the analysis. Before each question, a focus point appeared in the center of the screen for 400

TABLE 3
ABSOLUTE ESTIMATES (z-TRANSFORMED) BY ANCHOR, PLAUSIBILITY, AND TIME PRESSURE

	No time pressure		Time pressure	
Anchor	Plausible	Implausible	Plausible	Implausible
High	.09 ($N = 19$)	.11 ($N = 19$)	.08 ($N = 23$)	.48 ($N = 23$)
Low	−.02 ($N = 19$)	−.31 ($N = 19$)	.01 ($N = 23$)	−.62 ($N = 23$)

milliseconds, followed by the question, which remained on the screen until the first answer key was pressed. After a pause of 3 seconds, the next question was presented in the same sequence.

Preliminary analysis. Note that participants under time pressure, in contrast to those who were not under time pressure, did not necessarily have to answer the comparative question. Supplementary analysis indicated that in the time-pressure condition, 75% of the comparative questions were answered and all of the participants managed to answer at least half of the comparative questions. Thus, it appears reasonable to assume that participants under time pressure attempted to process the comparative questions but in some cases simply failed to do so within the time limits we imposed.

To make sure that under time pressure, task completion did not influence the critical response latencies to the absolute question, we compared latencies for participants who completed all of the four critical comparative questions ($N = 7$) with those who failed to complete at least one of them ($N = 15$).[5] Specifically, we conducted a preliminary 2 (all comparative tasks completed vs not all tasks completed) × 2 (high anchor vs low anchor) × 2 (plausible anchor vs implausible anchor) mixed model ANOVA, using the logarithmic transformations of response latencies to the absolute question as dependent variables. In this analysis, none of the effects involving Task Completion reached significance, $F(1, 20) = 1.22$, $p > .25$, for the main effect of Task Completion, $F(1, 20) < 1$, for all interaction effects. Thus, task completion did not influence response latencies to the absolute question and was not included in the main analysis.

Results

Absolute estimates. Table 3 shows mean *z*-transformed absolute judgments as a function of anchor (high vs low), plausibility and time pressure. High anchors led to generally higher estimates ($M = .19$) than low anchors ($M = -.24$), $F(1, 40) = 9.07$, $p < .01$. However, this difference was significantly more pronounced when anchors were implausible (.30 vs −.47) than when they were plausible (.09 vs −.01), $F(1, 40) = 5.34$, $p < .03$, for the interaction. This finding is in line with our earlier results (Strack & Mussweiler, 1997, Study 3) in which we demonstrated a tendency for stronger numeric assimilation of implausible anchors. More important is the effect of time pressure on the absolute estimates. As is apparent in Table 3, the effect of Anchor did not depend on Time Pressure, $F < 1$. In fact, no other effects were significant (all p's $> .2$).

Response latencies. Four participants were excluded because their latencies deviated from the mean by more than three standard deviations. Thus the analysis of the response latencies is based on the responses of 38 participants. As

[5] One participant was excluded from this analysis because his or her latency deviated from the mean by more than three standard deviations.

TABLE 4
RESPONSE LATENCIES FOR THE ABSOLUTE QUESTION BY ANCHOR, PLAUSIBILITY, AND TIME PRESSURE

	No time pressure		Time pressure	
Anchor	Plausible	Implausible	Plausible	Implausible
High	5167 ($N = 16$)	6063 ($N = 16$)	5884 ($N = 22$)	7259 ($N = 22$)
Low	4447 ($N = 16$)	6905 ($N = 16$)	6768 ($N = 22$)	9414 ($N = 22$)

Note. Response latencies are given in ms.

suggested by Fazio (1990), logarithmic transformations of response latencies were conducted to reduce the skewness of the response distribution. Our analyses are based on these logarithmic transformations. For ease of interpretation, however, we report the non-transformed means.

Because the pattern of response latencies for the *comparative task* is in part an artifact of the time pressure manipulation they are not reported here.

Response times to the *absolute question* are depicted in Table 4. These latencies were longer, when the comparative question had to be answered under time pressure ($M = 7331$ ms) than under no time pressure ($M = 5646$ ms), $F(1, 36) = 5.26$, $p < .03$, and when the anchor was implausible ($M = 7410$ ms) rather than plausible ($M = 5567$ ms), $F(1, 36) = 21.02$, $p < .001$. The effect of time pressure is not contingent on the plausibility of the anchor, $F(1, 36) < 1$, for the interaction.

Separate analysis for the no-time-pressure condition. The data obtained in the no time-pressure condition confirm the results of our earlier study (Strack & Mussweiler, 1997, Study 3). Specifically, response times in the comparative task were shorter for implausible anchors ($M = 4866$ ms) than for plausible anchors ($M = 6374$ ms), whereas the opposite is true in the absolute task (6484 ms vs 4807 ms), $F(1, 15) = 60.3$, $p < .001$, for the interaction.

Discussion

Study 3 demonstrates that absolute response latencies increase when the comparative question is answered under time pressure. Time pressure decreases the amount of knowledge that becomes accessible in the course of performing the comparative task and, therefore, increases the time required to perform the subsequent absolute judgment task. Interestingly, the effect of time pressure on absolute response latencies is not contingent on the plausibility of the anchor. The processing of implausible anchors requires less evidence than the processing of plausible ones and, therefore, also leads relatively less evidence to be accessible for performing the absolute task. Moreover, the effects of time pressure and plausibility on the amount of evidence activated by the comparative task appear to be additive. This suggests that even for comparative questions that include

implausible anchors and can thus be answered on the basis of little evidence, the amount of knowledge generated to answer this question is further limited by time pressure.[6]

Although time pressure to perform the comparative judgment task increased the time required to make absolute judgments, it did not influence the judgments themselves. This suggests that it is the *selectivity* of the evidence generated in order to solve the comparative task rather than its amount that leads to anchoring. That is, although time pressure to perform the comparative task restricted the amount of judgment-relevant information that was retrieved in the course of performing the task, it did not influence the type or implications of this evidence.

One additional aspect of our data seems noteworthy. Specifically, the anchoring effect was more pronounced for implausible than for plausible anchors. This result is consistent with previous research demonstrating that anchoring effects do also result for extremely implausible anchor values (Chapman & Johnson, 1994; Quattrone, Lawrence, Warren, Souza-Silva, Finkel, & Andrus, 1984; Strack & Mussweiler, 1997). From a selective accessibility perspective this finding seems surprising, because in order to compare the target to an implausible anchor value (e.g., Is the river Elbe longer or shorter than 45,000 kilometers?), judges only have to generate a minimal amount of knowledge about the target object (e.g., No river is 45,000 kilometers long) which would subsequently be accessible. As a result, one may expect anchoring to be less pronounced for implausible than for plausible anchors. Our data, demonstrate that this is not the case and thus suggest that the effects of plausible and implausible anchors may be mediated by different processes. We will further discuss these processes in the General Discussion.

In sum, the results of Study 3 support the notion that—at least for plausible anchors—anchoring is mediated by the use of semantic knowledge that has been rendered easily accessible in the course of solving the comparative task. Thus, in line with the Accessibility Hypothesis, anchoring appears to be mediated by a process that is akin to semantic priming. From this perspective, however, one of the most striking characteristics of the anchoring phenomenon—namely its extraordinary robustness—seems surprising.

Explaining the Robustness of Anchoring: A Self-Generation Effect

One of the most impressive demonstrations of this robustness stems from research on the impact of correctional instructions. Specifically, Wilson et al.

[6] In principle, one could argue that time pressure may have a direct effect on absolute response latencies so that the obtained findings are not mediated by the assumed knowledge generation process. For instance, participants may take longer to answer absolute questions in order to compensate for or recover from the time pressure exerted during the comparative judgment. Although our data cannot rule out this alternative interpretation for certain, a direct effect of time pressure on absolute response latencies seems unlikely as a mediator of our findings. First, our preliminary analysis demonstrates that task completion is not responsible for our findings. Second, the fact that the plausibility and the time pressure manipulation have parallel effects suggests that both are mediated by a similar mechanisms, namely the differential generation of knowledge that is relevant for the subsequent absolute judgment.

(1996) demonstrated that participants still showed the usual assimilation effect, if they were explicitly forewarned about the potential distortion. Such correctional instructions had no impact, even when participants were informed about the direction of the possible influence. In marked contrast to this extraordinary robustness, priming effects are rather fragile. Specifically, they are influenced by factors of which anchoring effects are independent, such as the extremity of the primes (Herr, 1986; Herr, Sherman, & Fazio, 1983), motivational factors (Martin, Seta, & Crelia, 1990), awareness of the influence (e.g., Strack, Schwarz, Bless, Kübler, & Wänke, 1993) and correctional instructions (e.g., Wegener & Petty, 1995). This striking disparity may be due to one important difference between the standard priming paradigm and the anchoring paradigm: In the standard priming paradigm, the primed information is usually *externally* provided by the experimenter, whereas in the anchoring paradigm, it is *internally* generated by the participants themselves. Recent research on belief perseverance (Davies, 1997) suggests that such *self-generation* may be responsible for the robustness of the anchoring effect. Specifically, it has been demonstrated that beliefs which are based on self-generated explanations are more persistent after evidential discrediting than those based on externally provided explanations (Davies, 1997). Thus, self-generation of explanations increased the robustness of beliefs. By the same token, self-generation of anchor-consistent evidence may produce exceptionally robust effects.

To recognize the role self-generation may play for the described disparity between anchoring and priming, it is important to specify the mechanisms that are responsible for the fragility of semantic priming effects. It has been suggested that priming effects are so fragile, because participants correct for influences that are not representative for their judgment (for a discussion of the role of representativeness in judgmental correction, see Strack, 1992). Such correction presupposes that judges are aware of the potentially contaminating influence (Strack, 1992; Wilson & Brekke, 1994). However, it seems difficult for people to recognize the invalidity of self-generated information (cf. Wilson, Hodges, & LaFleur, 1995). As a result, biased information may readily be used for a judgment, when it appears to be self generated. This possibility has been demonstrated in a study (Wilson et al., 1995) in which participants who first received biased information about a target person were either asked to recall this information before judging their liking for the target, or to analyze and list the reasons for their feelings about the target. If participants had to recall the provided information, judgments remained uninfluenced by it. If, however, participants had to list the reasons for their liking, their judgments were assimilated to the implications of the provided information. Thus, judges appear to have used this information as a basis for their judgment. This may be the case because information that comes to mind while searching for reasons may be seen as less biased than information that is recalled as having been presented beforehand. Similarly, solving the comparative anchoring task may not be recognized as a biasing determinant and the information that comes to mind while generating the absolute judgment may thus be attributed to

the target and not to an extraneous influence (see also Higgins, 1997). As a consequence, judges are unlikely to correct and may readily make use of easily accessible evidence.

Moreover, there is reason to believe that even if judges are aware of the contaminating influence of self-generated knowledge, they may fail to correct *sufficiently.* It has been suggested that the magnitude of correction corresponds to the perceived magnitude of the judgmental distortion (e.g., Strack, 1992; Wegener & Petty, 1995, 1997; Wilson & Brekke, 1994): The stronger the distorting influence is perceived to be, the more judges correct. Consequently, correction only compensates sufficiently for a distorting influence if judges are aware of the full amount to which their judgment is influenced. For externally provided material, judges typically seem to *over*estimate this magnitude, so that correction overcompensates and leads to contrast (e.g., Strack et al., 1993). Self-generated material, however, seems to be processed more deeply than externally provided material (Slamecka & Graf, 1978) and may thus have stronger effects on judgment. This assumption is also supported by research on knowledge accessibility effects which demonstrates that under specific circumstances, self-generated primes may influence judgment more strongly than externally provided primes (Smith & Branscombe, 1987, 1988). By the same token, generating knowledge during the comparative anchoring task may increase its accessibility to a larger extent and may thus produce stronger effects on the subsequent absolute judgment. In this situation, judges are likely to *under*estimate the amount to which their judgment is influenced, so that their corrective attempts may fall short of compensating for this influence completely. As a consequence, absolute estimates may be influenced by this knowledge, although judges tried to correct.

Thus, self-generation may contribute to the robustness of anchoring effects in at least two distinct ways. First, it may induce judges to see easily accessible evidence as representative for the target thus preventing judgmental correction. Second, it may produce stronger effects on the absolute judgment because it increases the accessibility of anchor-consistent knowledge to a larger extent, so that correction does not compensate completely for the influence.[7] Study 4 was designed to examine the role self-generation plays in anchoring.

STUDY 4

To explore in how far self-generation contributes to the robustness of the anchoring phenomenon, we combined the standard anchoring paradigm with a thought-listing procedure. In particular, the thoughts that came to participants' minds while working on the comparative task were listed after they had provided

[7] Clearly, some of the standard priming tasks such as unscrambling sentences (Srull & Wyer, 1979) also involve some activity on the side of the participants, so that here priming may also be seen as self generated. In contrast to the comparative anchoring task, however, here the implications of the provided evidence are clearly determined by the task itself (e.g., the provided words can only be arranged into a sentence with aggressive content). As a consequence, easily accessible evidence is more likely to be seen as contaminated.

their comparative judgment. Using a yoked-subjects design, these thoughts were either generated by the participants themselves, or by one of the other participants. Thus, experimental participants were either instructed to list the judgment-relevant thoughts that came to their mind while they had worked on the comparative task or they were instructed to read the thoughts that came to another participant's mind.

We expected that for self-generated thoughts the size of the anchoring effect would not differ from a control group that listed judgment-*ir*relevant thoughts. In line with the above reasoning, this should either be the case because self-generated thoughts are seen as representative for the target, so that no correction is carried out, or because self-generation yields stronger effects, so that correction is insufficient. In contrast, for externally provided thoughts, the anchoring effect should be smaller than in these two groups. Specifically, judges may see thoughts stemming from another participant as nonrepresentative for the target and thus as potentially contaminating their own judgment. Consequently, they may try to correct for their influence which attenuates the effect.

Method

Participants. We recruited 90 male and female non-psychology students of the University of Würzburg as participants. They were asked to take part in a pretest for the construction of a questionnaire assessing general knowledge and were offered a chocolate bar as compensation.

Materials. The questionnaire consisted of two pairs of general knowledge questions (''annual mean temperature in the Antarctic'' and ''length of the River Elbe'') similar to those used in the last study. The high and low plausible values depicted in Table 2 were used as anchors. One question included a high anchor, the other one included a low anchor. To control for content and order effects, two versions of the questionnaire were constructed. In both versions, the question pertaining to the Antarctic preceded the one pertaining to the Elbe. However, the anchor condition was assigned to different questions in the two versions. Specifically, in version one the first question included the high anchor, whereas the second question included the low anchor. For version two this assignment was reversed.

Procedure. Participants took part in the experiment in groups of up to 10. They were recruited in the university cafeteria and led to a separate room where they received the questionnaire. The instructions were similar to those used in Studies 1 and 2. However, they included one additional paragraph which explained the purpose of the thought-listing procedure. Specifically, participants were told that although general knowledge questionnaires were widely used, little was known about how people answer such questionnaires. To investigate the strategies used, we would also analyze what people thought of while answering the questions. One third of the participants were further instructed that therefore, we would ask them to list the features of the target that came to their mind while answering two of the questions. The second third was told that we would present them a list of features that came to another person's mind while he or she answered the questions. The remaining third of control participants were informed that we would ask them to list the thoughts that came to their mind although they did not pertain to the target.

The general knowledge questions succeeded the instructions. Following both comparative questions, participants received one of three thought-listing instructions. Specifically, they were either instructed to list those features of the target that came to their mind while solving the preceding task (i.e., judgment-relevant thoughts), received a list of judgment-relevant thoughts that was generated by one of the other participants, or they were instructed to list the thoughts that came to their mind while working on the comparative task although they did not pertain to the target (i.e., judgment-irrelevant thoughts). Thus, for both question pairs, participants first answered the comparative question, then worked on the list of features and finally answered the absolute question.

Results

We excluded one participant from the analysis, because he did not answer all of the comparative questions. Two additional participants were excluded, because one of their absolute estimates deviated from the question mean by more than 3 standard deviations. Thus, the analysis is based on the responses of the remaining 87 participants.

Absolute estimates. As in the previous studies, the *z*-transformed answers to the absolute question served as the central dependent variables. Overall, high anchors led to higher estimates ($M = .32$) than low anchors ($M = -.38$), $F(1, 84) = 30.23$, $p < .001$. Thus, the typical anchoring effect was replicated. More interesting, the magnitude of this effect depended on the thought-listing condition. Specifically, differences between estimates for the high and the low anchor (i.e., the anchoring effect) were more pronounced for control participants (.44 vs −.47) and participants who generated features of the target themselves (.30 vs −.67) than for those who received features generated by another participant (.22 vs .02), $F(2, 84) = 3.67$, $p < .03$, for the interaction. Contrast analyses revealed that the magnitude of the anchoring effect did not differ for the first two groups, $t(84) = .23$, $p > .8$. Compared to these two groups, however, the amount of anchoring was significantly smaller in the group that received the thought list generated by another participant, $t(84) = 2.52$, $p < .02$.

Thought content. To analyse the content of the thought lists, two independent judges who were blind to experimental conditions rated the implications of the listed thoughts for the extension of the target. They used 5-point rating scales that ranged form −2 (e.g., ''indicating very short extensions of the Elbe'') to +2 (e.g., ''indicating very long extensions of the Elbe'') to do so. Both judges showed high agreement ($r = .8$, for the Antarctic and $r = .81$ for the Elbe) so that their ratings were combined into one single score for each target.

In the irrelevant thought-listing condition,[8] the implications of the listed thoughts were independent of the presented anchor value ($M = -.15$, for the high anchor, and $M = -.13$, for the low anchor). In the relevant thought-listing condition, however, participants were more likely to list thoughts that implied high values for the target (e.g., ''the Elbe is one of the longest rivers in Europe'') when they had previously considered the high anchor ($M = .26$), and were more likely to list thoughts that implied low values (e.g., ''the Antarctic has the lowest temperatures on earth'') when they had compared the target to the low anchor ($M = -.69$), $F(1, 57) = 4.86$, $p < .03$, for the interaction of Thought Listing and Anchor.

Discussion

In line with our hypothesis, these results suggest that self-generation contributes to the robustness of the anchoring effect. The magnitude of the anchoring

[8] Examples of the irrelevant thoughts that were listed, include: ''the experimenter has her finger nails painted black,'' ''weird question,'' ''I'm in a hurry,'' etc.

effect for participants who generated judgment-relevant features of the target themselves did not differ from that of the control group. In contrast, less anchoring was evident in the estimates of participants who were exposed to features generated by another participant. Notably, because Study 4 used a yoked design, these differences cannot be mediated by the content of the listed thoughts. However, our findings seem consistent with the assumption that participants corrected for the influence of the externally-provided list, which led to an attenuation of the effect. This interpretation of our data is also in line with current conceptualizations (e.g., Higgins, 1996; Martin & Achee, 1992; Schwarz & Bless, 1992; Strack, 1992) which assume that an attenuation of priming effects is typically caused by attempts to correct for unwanted influences on judgment. It is important to note, however, that Study 4 explicitly focussed on the role self-generation plays in the robustness of anchoring effects rather than the mechanisms that mediate the effects of self-generation. As a consequence, our results do not provide direct support for the assumed correction mechanism. Thus, the exact processes that are responsible for the robustness of self-generated knowledge remain to be investigated.

Most important, however, the results of Study 4 help resolve the seeming contradiction that anchoring effects are remarkably robust, although they appear to be mediated by a mechanism that is akin to the fragile semantic priming mechanism. In the anchoring paradigm, participants generate judgment-relevant knowledge themselves which may increase the robustness of the effect.

Moreover, the analysis of the listed thoughts provides additional support for the Selective Accessibility model. Consistent with the model's fundamental assumption, participants primarily listed anchor-consistent thoughts. This suggests that solving a comparative anchoring task, in fact, selectively increased the accessibility of anchor-consistent knowledge (for further evidence, see Mussweiler & Strack, 1998).

GENERAL DISCUSSION

Taken together, the present findings are consistent with the assumption that anchoring is mediated by mechanisms of selective accessibility. In particular, participants appear to solve the standard comparative task by testing the possibility that the target's extension along the judgmental dimension is equal to the anchor value (cf. Studies 1 and 2). In performing this test, they seem to apply a hypothesis-consistent test strategy. That is, judges selectively generate semantic knowledge that is consistent with the notion that the target's value is equal to the anchor value (cf. Study 4). Furthermore, to solve the subsequent absolute task, participants appear to resort to semantic knowledge that has been rendered easily accessible in the course of solving the comparative task (cf. Study 3). Because this evidence was selectively generated to be anchor consistent, its use leads to absolute estimates that are anchor-consistent as well. Thus, anchoring effects in the standard paradigm appear to be mediated by the joint influence of hypothesis-consistent testing and semantic priming. Finally, our data hint at the mechanism

that is responsible for the remarkable robustness of the anchoring phenomenon. Specifically, comparing the target to the anchor value may yield such a robust influence on absolute estimates, because the evidence used to make the comparative judgment is self generated (cf. Study 4).

Beyond the Standard Anchoring Paradigm: Implicit Comparison Processes

In the present framework, anchoring effects are obtained because comparing a target to an anchor value selectively increases the accessibility of anchor-consistent knowledge about the target. This mechanism presupposes that the target is compared to the anchor value which is a requirement in the anchoring paradigm used in the current research. Anchoring effects, however, have also been obtained when no explicit comparison is required. For example, Northcraft and Neale (1987) demonstrated that merely presenting the listing-price of a house influenced participants' estimates of its value. In fact, even completely irrelevant numbers may exert an effect if their accessibility is sufficiently increased in a preceding unrelated task (Wilson et al., 1996).

Although at first sight, these findings seem to contradict the proposed model, it seems plausible to assume that to generate an absolute estimate, participants may select a standard of comparison themselves against which they evaluate the target. This assumption that in a situation of judgmental uncertainty, an absolute dimensional judgment constitutes an implicit comparative judgment is consistent with a number of classic approaches which hold that human judgment is essentially comparative in nature (e.g., Festinger, 1954; Helson, 1964; Kahneman & Miller, 1986). For example, Kahneman and Miller's (1986) norm theory proposes that events of all levels of complexity are evaluated in comparison to an evoked norm. Moreover, in the presence of a specific standard such comparisons may arise spontaneously (Gilbert, Giesler, & Morris, 1995). Similarly, when no explicit comparison is required, absolute judgments may involve an implicit comparison of the target with a salient standard.

At least two mechanisms may guide the selection of such a standard. *Conversational inferences* (Grice, 1975) constitute a first possibility. For example, participants in Northcraft and Neale's (1987) study may well have inferred that the provided listing-price is relevant for the judgment to be made and thus used it as an initial standard of comparison. Alternatively, the selection of a standard may be guided by its *accessibility* (for a review, see Higgins, 1996). Thus, anchor values that are highly accessible due to their extensive use in a preceding task (cf. Wilson et al., 1996) may exert an effect because they are implicitly used as standards of comparison.

After a standard of comparison is selected, it is compared to the target object. This comparison process requires the generation of semantic knowledge about the target and is thus likely to involve the described mechanisms of selective accessibility. From this perspective, judgmental anchoring is a *two-stage process:* In a first judgment stage, judges have to select an appropriate standard of comparison. This initial *selection of an anchor* may be guided by conversational

inferences or the accessibility principle. The subsequent *comparison of the target with this anchor,* however, is likely to involve mechanisms of selective accessibility. In light of this analysis, there exists reason to believe that the described selective accessibility mechanism is not restricted to the standard anchoring paradigm. Rather, it may also contribute to anchoring effects which involve different judgmental tasks (e.g., Northcraft & Neale, 1987; Wilson et al., 1996). Clearly, the supposed generality of the selective accessibility mechanism—although plausible on theoretical grounds—remains to be tested.

Explaining the Effects of Implausible Anchors: Adjustment and Selective Accessibility

This conceptualization of anchoring as a two-stage process may also help explain the effects of implausible anchor values. As noted above, our data suggest that the effects of implausible and plausible anchors are mediated by different processes. First, implausible anchors yield stronger assimilation effects than plausible anchors. Second, response latencies for the comparative and the absolute task differ for plausible and implausible anchors. Specifically, comparative response latencies are shorter for implausible than for plausible anchors, while absolute response latencies are longer. The fact that the comparative question is answered faster suggests that less semantic knowledge is generated for implausible than for plausible anchors. In fact, to answer comparative questions that include implausible anchors, participants do not necessarily have to generate evidence pertaining to the specific object of the task (i.e., *individuating knowledge*). Instead, it is sufficient to generate knowledge about the superordinate category of the target (i.e., *categorical knowledge*) (for a more detailed discussion of this possibility, see Mussweiler & Strack, 1998).[9] For example, participants asked whether the river Elbe is longer or shorter than 45,000 kilometers do not have to generate information about the river Elbe in particular. Rather, it is sufficient to recall that rivers in general are shorter than 45,000 kilometers. Thus, testing the focal hypothesis of a comparative question that includes an implausible anchor does not necessarily require the selective generation of anchor-consistent evidence, as suggested by the Selective Accessibility Model. Consequently, the fact that implausible anchors lead to anchoring effects nevertheless calls for an alternative explanation.

A combination of an anchoring-and-adjustment rationale (Tversky & Kahneman, 1974; Quattrone et al., 1984) and a selective accessibility mechanism constitutes a promising candidate in the quest for an adequate explanation. In particular, participants may process implausible anchors by first adjusting to the boundary value of a distribution of plausible values and then testing the hypoth-

[9] This assumption that participants resort to categorical knowledge if it is sufficient to make an adequate judgment is also consistent with the literature on information use in impression formation (cf. Brewer, 1988; Fiske & Neuberg, 1990). Here, it has been argued that people rely more heavily on categorical knowledge, if doing so yields a judgment that is sufficiently adequate with respect to their involvement in the judgment and the fit between individuating and categorical information.

esis that the object's extension is equal to this boundary value. That is, they may first select an appropriate standard of comparison by adjusting from the provided inappropriate (i.e. implausible) value and then compare the target object to this selected standard. In our example, they may start with the implausible anchor of 45,000 kilometers, adjust until the first plausible value is reached (e.g., 1300 km), and then test the hypothesis that the length of the river Elbe is equal to this value. Thus, the boundary value may serve as an anchor.

This rationale is able to account for all of the results obtained for implausible anchors. First, the amount of anchoring is larger for implausible anchors, because in most cases the boundary value used to generated semantic knowledge will be more extreme than the plausible anchors provided. Second, response latencies to the comparative task are faster, because solving the task requires the generation of less individuating knowledge. Finally, response latencies to the absolute task are longer, because the accessibility of relevant knowledge has been increased to a lesser degree in the course of solving the comparative task.

Thus, selective accessibility may not be limited to plausible anchor values. Rather, in combination with the described anchoring-and-adjustment process (Tversky & Kahneman, 1974), it may also contribute to the effects of implausibly extreme anchors.

Anchoring, Knowledge, and Uncertainty

Conceivably, the described selective accessibility mechanism critically depends on the knowledge participants have about the target object. That is, assuming that anchoring is mediated by a selective search of judges' knowledge base about the target implies that the content of this knowledge base is a critical variable in the anchoring process. In order to specify the role knowledge may play in anchoring, we (Mussweiler & Strack, 1998) have recently suggested that judges' knowledge about the target may be characterized by a distribution of possible values for the target (cf. Wyer, 1973). The dispersion of this distribution then depends on the amount of knowledge a particular judge has about the judgmental target. The more a judge knows (i.e., the less uncertain he or she is), the narrower his or her range of plausible values. Specifically, the distribution of possible values will be extremely narrow, if a judge has maximal knowledge about the target (e.g., because he or she knows the exact value), it will be extremely wide, if a judge has minimal knowledge about the target (e.g., because he or she only knows to which general category the target belongs).

In this conceptual framework, the dispersion of the described distribution determines the plausibility of a provided anchor value. Specifically, any anchor value that lies within the boundaries of the probability distribution constitutes an acceptable and plausible value for the target. Any anchor that lies outside of these boundaries, however, constitutes an unacceptable and implausible value. As suggested before, comparisons with plausible and implausible anchor values are likely to involve different mechanisms. Whereas a plausible anchor value appears to be directly used to generate anchor-consistent information, an implausible

anchor value may only be used as a starting-point to determine an acceptable value, which may then serve as a standard for the selective accessibility mechanism.

Thus, depending on the amount of knowledge judges have about the target object the very same anchor value may be processed quite differently. These different processes may ultimately yield different estimates for the target. For example, consider a judge who knows that the river Elbe is a European river and that it is shorter then the river Rhine. For this judge an anchor value of 2000 kilometers is likely to be implausible because it exceeds the actual length of the Rhine (1320 kilometers). Consequently, he or she is unlikely to directly use the provided value for the selective accessibility mechanism. Rather, he may use the first acceptable value (e.g., 1300 kilometers) as a standard and generate evidence that is consistent with the assumption that the river Elbe is 1300 kilometers long. In contrast, the very same anchor value of 2000 kilometers may be plausible for a judge who only knows that the river Elbe is a European river. Consequently, this judge may test the hypothesis that the river Elbe is indeed 2000 kilometers long by generating evidence that is consistent with this assumption. Because the subsequent absolute judgment is then based on the implications of the generated knowledge, it is likely to be higher for the latter judge. Thus, the judge with little knowledge about the target is likely to be more strongly influenced by the provided anchor value (i.e. show more anchoring).

In line with the predictions derived form this conceptualization, we (Mussweiler & Strack, 1998) have recently demonstrated that the size of the anchoring effect depends on the amount of knowledge judges have about the target object: The more judges know about the judgmental target, the less they are influenced by a provided anchor value (see also Jacowitz & Kahneman, 1995; Wilson et al., 1996). Moreover, the current conceptualization implies that the influence that judges' knowledge has on the size of the anchoring effect may be mediated by differences in the perceived plausibility of the anchor value. Specifically, judges with little knowledge about the target object may be more susceptible to anchoring because they are likely to perceive even extreme anchor values as plausible and test the hypothesis that the target value is equal to this value.

CONCLUSION

In the present paper we have drawn on the general principles of hypothesis-consistent testing and semantic priming to explain the anchoring effect. The implications of our analysis, however, are not limited to the anchoring paradigm. Rather, conceptualizing judgmental anchoring as a hypothesis-testing phenomenon, allows us to explore the significance of our findings for processes of hypothesis-testing in general. From this perspective, one striking characteristic of the anchoring phenomenon is that using what appears to be the most efficient strategy to solve the comparative task (i.e., hypothesis-consistent testing) distorts the subsequent absolute judgment. By the same token, strategies of hypothesis-consistent testing that are believed to allow for the most critical test of a given

hypothesis in many situations (see Klayman & Ha, 1987; Trope & Liberman, 1996), may also have some costs that are delayed. Specifically, although applying such a strategy may yield an adequate initial judgment, it may also prepare the ground for a distortion of subsequent judgments.

REFERENCES

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition* (pp. 1–36). Hillsdale, NJ: Erlbaum.

Carlson, B. W. (1990). Anchoring and adjustment in judgments under risk. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **16,** 665–676.

Cervone, D., & Peake, P. K. (1986). Anchoring, efficacy, and action: The influence of judgmental heuristics on self-efficacy judgment and behavior. *Journal of Personality and Social Psychology,* **50,** 492–501.

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

Clark, H. H., & Schober, M. F. (1992). Asking questions and influencing answers. In J. M. Tanur (Ed.), *Questions about questions. Inquiries into the cognitive bases of surveys* (pp. 15–47). New York: Russel Sage Foundation.

Cohen, J. (1977). *Statistical power analysis for the social sciences.* New York: Academic Press.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *The Behavioral and Brain Sciences,* **4,** 317–331.

Davies, M. F. (1997). Belief persistence after evidential discrediting: The impact of generated versus provided explanations on the likelihood of discredited outcomes. *Journal of Experimental Social Psychology,* **33,** 561–578.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology,* **32,** 53–88.

Fazio, R. H. (1990). A practical guide to the use of response latency in social psychological research. In C. Hendrick & M. S. Clark (Eds.), *Research methods in personality and social psychology* (pp. 74–97). Newbury Park, CA: Sage.

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

Fischhoff, B. (1975). Hindsight = foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology,* **89,** 288–299.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 1–74). New York, NY: Academic Press.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review,* **103,** 592–596.

Gilbert, D. T., Giesler, R. B., & Morris, K. A. (1995). When comparisons arise. *Journal of Personality and Social Psychology,* **69,** 227–236.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York: Academic Press.

Hasher, L., Attig, M. S., & Alba, (1981). I knew it all along: Or did I? *Journal of Verbal Learning and Verbal Behavior,* **20,** 86–96.

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin,* **107,** 311–327.

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

Herr, P. M. (1986). Consequences of priming: Judgment and behavior. *Journal of Personality and Social Psychology,* **51,** 1106–1115.

Herr, P. M., Sherman, S. J., & Fazio, R. H. (1983). On the consequences of priming: Assimilation and contrast effects. *Journal of Experimental Social Psychology,* **19,** 323–340.

Higgins, E. T. (1989). Knowledge accessibility and activation: Subjectivity and suffering from unconscious sources. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 75–123). New York: Guilford.

Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133–168). New York: Guilford.

Higgins, E. T. (1997). Biases in social cognition: ''Aboutness'' as a general principle. In C. McGarty & S. A. Haslam (Eds.), *The message of social psychology: Perspectives on mind in society* (pp. 183–199). Cambridge, MA: Blackwell.

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology,* **13,** 141–154.

Holtgraves, T., & Skeel, J. (1992). Cognitive biases in playing the lottery: Estimating the odds and choosing the numbers. *Journal of Applied Social Psychology,* **22,** 934–952.

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

Johnson, E. J., & Schkade, D. A. (1989). Bias in utility assessments: Further evidence and explanations. *Management Science,* **35,** 406–424.

Johnson-Laird, P. N. (1983). *Mental models: Toward a cognitive science of language, inference, and consciousness.* Cambridge, MA: Harvard University Press.

Joyce, E. J., & Biddle, G. C. (1981). Anchoring and adjustment in probabilistic inferences in auditing. *Journal of Accounting Research,* **19,** 120–145.

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

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypotheses testing. *Psychological Review,* **94,** 211–228.

Kruglanski, A. W., & Webster, D. M. (1996). Motivated closing of the mind: ''Seizing'' and ''freezing.'' *Psychological Review,* **103,** 263–283.

Leyens, J.-P., Yzerbyt, V., & Corneille, O. (1996). The role of applicability in the emergence of the overattribution bias. *Journal of Personality and Social Psychology,* **70,** 219–229.

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

Lopes, L. (1991). The rhetoric of irrationality. *Theory and Psychology,* **1,** 65–82.

Martin, L. L., Seta, J. J., & Crelia, R. A. (1990). Assimilation and contrast as a function of people's willingness and ability to expend effort in forming an impression. *Journal of Personality and Social Psychology,* **59,** 27–37.

Martin, L. L., & Achee, J. W. (1992). Beyond accessibility: The role of processing objectives in judgment. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgment* (pp. 249–275). Hillsdale, NJ: Erlbaum.

Mayseless, O., & Kruglanski, A. W. (1987). What makes you so sure? Effects of epistemic motivations on judgmental confidence. *Organizational Behavior and Human Decision Processes,* **39,** 162–183.

Mussweiler, T., Fō rster, J., & Strack, F. (1997). Der Aukereffekt in Abhā ngigkeit von der Anwenelbarkeit ankerkonsistenter Information: Ein Modell Selektive Zugā nglichkeit. *Zeitschrift fū r Experimentelle Psychologie,* **44,** 589–615.

Mussweiler, T., & Strack, F. (1998). *The use of category and exemplar knowledge in the solution of anchoring tasks.* Manuscript submitted for publication, University of Würzburg.

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

Northcraft, G. B., & Neale, M. A. (1987). Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes,* **39,** 84–97.

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

Pohl, R. F. (1996). Der Rückschaufehler—Eine systematische Verfälschung der Erinnerung [The hindsight bias—A systematic memory distortion]. *Report Psychologie,* **21,** 596–609.

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., Warren, D. L., Souza-Silva, K., Finkel, S. E., & Andrus, D. E. (1984). *Explorations in anchoring: The effects of prior range, anchor extremity, and suggestive hints.* Unpublished manuscript.

Rosenthal, R., & Rosnow, R. L. (1985). *Contrast analysis: Focused comparisons in the analysis of variance*. New York: Cambridge University Press.

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

Schwarz, N. (1994). Judgment in social context: Biases, shortcomings, and the logic of conversation. In M. P. Zanna (Ed.), *Advances in experimental social psychology.* (pp. 125–162). San Diego, CA: Academic Press.

Schwarz, N., & Bless, H. (1992). Constructing reality and its alternatives: An inclusion/exclusion model of assimilation and contrast effects in social judgment. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgment* (pp. 249–275). Hillsdale, NJ: Erlbaum.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology,* **61,** 195–202.

Sedikides, C., & Skowronski, J. J. (1991). The law of cognitive structure activation. *Psychological Inquiry,* **2,** 169–184.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory,* **4,** 592–604.

Smith, E. R., & Branscombe, N. R. (1987). Procedurally mediated social inferences: The case of category accessibility effects. *Journal of Experimental Social Psychology,* **23,** 361–382.

Smith, E. R., & Branscombe, N. R. (1988). Category accessibility as implicit memory. *Journal of Experimental Social Psychology,* **24,** 490–504.

Snyder, M. L., & Swann, W. B. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology,* **33,** 1202–1212.

Srull, T. K., & Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology,* **37,** 1660–1672.

Strack, F. (1992). The different routes to social judgments: Experiential versus informational strategies. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgment* (pp. 249–275). Hillsdale, NJ: Erlbaum.

Strack, F., Erber, R., & Wicklund, R. A. (1982). Effects of salience and time pressure on ratings of social causality. *Journal of Experimental Social Psychology,* **18,** 581–594.

Strack, F., & Martin, L. L. (1987). Thinking, judging, and communicating: A process account of context effects in attitude surveys. In H. J. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology.* (pp. 123–148). New York: Springer.

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

Strack, F., Schwarz, N., Bless, H., Kübler, A., & Wänke, M. (1993). Awareness of the influence as a determinant of assimilation versus contrast. *European Journal of Social Psychology,* **23,** 53–62.

Switzer, F. S., & Sniezek, J. A. (1991). Judgment processes in motivation: Anchoring and adjustment effects on judgment and behavior. *Organizational Behavior and Human Decision Processes,* **49,** 208–229.

Trope, Y., & Bassok, M. (1982). Confirmatory and diagnostic strategies in social information gathering. *Journal of Personality and Social Psychology,* **43,** 22–34.

Trope, Y., & Liberman, A. (1996). Social hypothesis testing: Cognitive and motivational factors. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 239–270). New York: Guilford.

Tversky, A. (1977). Features of similarity. *Psychological Review,* **84,** 327–352.

Tversky, A., & Gati, I. (1978). Studies of similarity. In E. Rosen & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 79–97). Hillsdale, NJ: Erlbaum.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology,* **5,** 207–232.

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

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology,* **12,** 129–140.

Wason, P. C., & Johnson-Laird, P. N. (1972).*Psychology of reasoning: Structure and content.* Cambridge, MA: Harvard University Press.

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

Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naive theories of bias in bias correction. In M. Zanna (Ed.), *Advances in Experimental Social Psychology.* San Diago, CA: Academic Press.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: unwanted influences on judgments and evaluations. *Psychological Bulletin,* **116,** 117–142.

Wilson, T. D., Hodges, S. D., & LaFleur, S. J. (1995). Effects of introspecting about reasons: Inferring attitudes from accessible thoughts. *Journal of Personality and Social Psychology,* **69,** 16–28.

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.

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

Wyer, R. S. (1973). Category ratings as ''subjective expected values'': Implications for attitude formation and change. *Psychological Review,* **80,** 446–467.

Wyer, R. S., & Srull, T. K. (1989). *Memory and cognition in its social context.* Hillsdale, NJ: Erlbaum.