

Anchoring effects in the judgment of confidence: Semantic or numeric priming?

STEVEN R. CARROLL, WILLIAM M. PETRUSIC, AND CRAIG LETH-STEENSEN
Carleton University, Ottawa, Ontario, Canada

Over the last decade, researchers have debated whether anchoring effects are the result of semantic or numeric priming. The present study tested both hypotheses. In four experiments involving a sensory detection task, participants first made a relative confidence judgment by deciding whether they were more or less confident than an anchor value in the correctness of their decision. Subsequently, they expressed an absolute level of confidence. In two of these experiments, the relative confidence anchor values represented the midpoints between the absolute confidence scale values, which were either explicitly numeric or semantic, nonnumeric representations of magnitude. In two other experiments, the anchor values were drawn from a scale modally different from that used to express the absolute confidence (i.e., nonnumeric and numeric, respectively, or vice versa). Regardless of the nature of the anchors, the mean confidence ratings revealed anchoring effects only when the relative and absolute confidence values were drawn from identical scales. Together, the results of these four experiments limit the conditions under which both numeric and semantic priming would be expected to lead to anchoring effects.

The study of cognitive phenomena has revealed few effects as reliable and omnipresent as the anchoring effect. Anchoring effects occur when decision makers begin to estimate an unknown but absolute quantity by first assessing the plausibility of some initially considered value. The decision maker adjusts and reconsiders this value until it finally becomes a reasonable estimate of the absolute quantity. Since the process of adjustment is often insufficient, the magnitude of the final estimate becomes “anchored” by the magnitude of the originally considered value. This effect was most famously illustrated by Tversky and Kahneman (1974). Participants in that study were first asked to decide whether the number of African countries in the United Nations (U.N.) was greater or less than a “random” value (either 10 or 65) before providing the researchers with an absolute judgment of the number of African countries in the U.N. These researchers found that the relative judgment value anchored the mean absolute judgment values, in that participants who considered the relative value 10 made reliably lower mean absolute judgments than did those who considered the relative value 65.

Researchers have since become divided on the causes and limitations of anchoring effects. Although it has been suggested that anchors exert their influence by priming the features of the subsequently selected absolute response (Chapman & Johnson, 1994; see also Jacowitz & Kahneman, 1995), a particular debate has arisen regarding whether these effects are best described as examples of numeric or semantic priming. Advocates of the numeric priming explanation suggest that any number considered in the relative judgment phase of an anchoring experiment

will serve to anchor the subsequent absolute judgment. Advocates of the semantic priming explanation suggest that anchoring effects will be strongest when the anchor considered by the decision maker is semantically comparable to the subject of the absolute judgment.

Strack and Mussweiler (1997) are among those who believe that anchoring effects are primarily the result of semantic priming. In one of their studies, participants were asked to judge either the width or the length of a bridge relative to an anchor before being subsequently asked to provide an absolute estimate of the length of the bridge. They found that anchoring effects were reduced when the relative and absolute judgments referred to different dimensions of the target. In a second study, Strack and Mussweiler demonstrated how unlikely anchors can produce contrast effects. For example, when participants were asked to decide whether the mean antarctic temperature was more or less than either -20°C (the high-anchor condition) or -50°C (the low-anchor condition) before estimating a mean temperature for Hawaii, anchoring effects were reversed (i.e., the estimates were slightly lower for the -20°C anchor than for the -50°C anchor). Taken together, these studies demonstrate how semantically relevant anchors exert more influence on estimates of magnitude than do semantically irrelevant anchors.

Alternatively, Wong and Kwong (2000) presented findings that support the numeric priming side of the anchoring debate. In one study, participants were first asked to decide whether a particular airport runway was longer or shorter than either 7.3 km (the small-anchor condition) or 7,300 m (the large-anchor condition). Subsequently, they

S. R. Carroll, srcarrol@connect.carleton.ca

were asked to estimate the cost of a new bus. As such, the anchors used in this experiment were both semantically identical to one another and semantically irrelevant to the target of the absolute estimation judgment. These researchers found that participants presented with the 7.3-km anchor made reliably lower estimates of the cost of the bus than did participants presented with the 7,300-m anchor. In a second study, Wong and Kwong asked participants to decide whether the mean winter temperature in Hong Kong was greater or less than 30°F, 1°F, -1°C, or -17°C. Subsequently, participants were asked to estimate the mean winter temperature in either Hong Kong or Singapore. These researchers found that it was the absolute (i.e., unsigned) value of the anchor used in the relative decision that exerted a reliable influence on the mean winter temperature estimates and, furthermore, that very similar anchoring effects occurred regardless of the congruity between the targets of both kinds of judgments. Contrary to the findings of Strack and Mussweiler (1997), Wong and Kwong concluded that anchoring effects are superficial in nature and suggested that although semantic congruity may enhance an anchoring effect, it does not cause it.

In a response to these findings, Mussweiler and Strack (2001) had participants in one experiment make a relative judgment regarding the length of an airport runway and then an absolute judgment of either the cost of a bus (as in Wong & Kwong, 2000) or the length of the runway. They found that the size of the observed anchoring effect was much larger when the relative judgment was semantically comparable to the absolute judgment. On the basis of this finding, as well as of those from two other experiments that they ran, these researchers drew a conclusion opposite to that of Wong and Kwong—namely, that numeric priming occurs only when the anchor and target are not semantically congruous.

Also of relevance here is work by Chapman and Johnson (1994), who demonstrated through a series of experiments that anchoring effects occur only if the relative and absolute judgments involve the same scale. For instance, in their second experiment, they showed that object anchors, such as *20 extra hours of sleep per week* or *a 10-year supply of Blatz beer*, did not have a subsequent influence on the absolutely expressed selling price of evaluated lotteries. Chapman and Johnson suggested that the lack of an anchoring effect was possibly due to the semantic incompatibility of the relative and absolute judgment scales. Namely, their participants may not have been able to equate extra sleep or free beer with money.

The present study serves to extend the findings of Chapman and Johnson (1994) by having participants provide both relative and absolute judgments of confidence (concerning their responses to a sensory-detection task) using either the same scales or different scales, while simultaneously evaluating the saliency of both numeric and semantic anchors. In a typical decision-making-with-confidence study, participants are asked to compare the physical properties of two stimuli and then to express their confidence in having just made a correct decision. Confidence is typically expressed via the use of either a scale involving semantic representations of magnitude (e.g., a linearly arranged series of labels

such as *guess, low, moderate, high, or certain*) or a numeric scale (e.g., 50–100, where participants are told that a confidence rating of 50 is indicative of a guess and 100 is indicative of certainty, with the values of 60, 70, 80, and 90 representing intermediate feelings of certainty).

In Experiment 1, we tested whether classic anchoring effects can be obtained for confidence judgments when participants are asked to make both relative and absolute judgments of confidence using numeric scale values. In Experiment 2, participants were asked to make both relative and absolute confidence judgments using alphabetic values. Such conditions provide a fairly strong test of whether anchoring effects are necessarily numerically based. In Experiments 3 and 4, relative and absolute judgments of confidence were made using different scales (i.e., semantic then numeric, and vice versa) that are, however, still semantically comparable (i.e., *guess* = 50, *certain* = 100). Such conditions allow for a determination of the extent to which semantic comparability alone is sufficient for anchoring effects to occur.

GENERAL METHOD

Participants

Eighty first-year undergraduate psychology students from Carleton University participated in this study in return for course credit. Twenty participants were assigned to each of the four presented experiments.

Apparatus

Each of the four experiments was conducted using a desktop computer equipped with a Pentium-class processor, a standard color monitor, and a Windows 98 operating system. Stimulus presentation and response data collection were controlled via Superlab Pro version 2.0. Participant responses were made via a control panel with two primary response buttons (labeled YES and NO), two secondary response buttons (labeled "<" and ">"), and seven confidence response buttons. The labels assigned to the confidence response buttons varied among the four experiments.

Stimuli

The stimuli used in these experiments consisted of sixteen 100 × 100 pixel squares arranged to form a larger 4 × 4 square. Each of the 16 squares contained two thousand five hundred 2 × 2 pixel dots. Each dot was either inactive (colored white) or active (colored black). The density level of the dots in the four center squares varied from trial to trial, but on any given trial, all of the center squares were homogeneously dense, with either 50% (noise trials) or 52%, 54%, 56%, or 58% (signal trials) active dots. Only 50% of the dots in the 12 outer background squares were active on any given trial.

Procedure

Participants were asked, on each trial, to decide whether the density of active (black) dots was greater in the center than around the edges of the display. Half of the trials in each block were noise trials (for which the correct response was *no*), and the other half were signal trials (for which the correct response was *yes*). Each of the four signal strengths occurred equally often within the signal trials in each of four blocks of 200 trials. One of the following two conditions was performed in the first two blocks of trials, whereas the other was performed in the second two blocks. The order in which these two conditions were performed was counterbalanced across participants.

In one condition, immediately following each signal detection decision, participants were asked to rate their confidence in the

accuracy of that decision (i.e., to provide an absolute confidence judgment). The type of scale used to make these confidence ratings varied depending on the experiment. In a second condition, following each signal detection decision but before they rated confidence, participants were first asked to decide whether their eventual confidence rating would be more or less than a confidence label presented on the computer monitor (i.e., to provide a relative confidence judgment). They were told to press the "<" button if they felt that they were less confident than this label but to press the ">" button if they felt that they were more confident than this label. As such, the label used for this relative confidence judgment served as a potential anchor for the subsequent absolute confidence evaluations. The value of these anchors varied depending on the experiment (see the following individual experiment descriptions). Within each experiment, each anchor value was presented an equal number of times at each level of signal strength within signal trials, as well as equal numbers of times for signal and noise trials.

Analyses

For each experiment, response times (RTs) less than 200 msec in duration (i.e., reflexive responses), as well as RTs more than 3 *SDs* above each participant's mean primary decision RT (i.e., distracted responses), were censored, as were the data for trials in which participants expressed a certainty of having made a mistake. Censored trials never accounted for more than 3.52% of the observed data. Only the results of those analyses relevant to the present discussion are presented here. More specifically, note that analyses conducted on the blocks of trials that did not involve relative confidence judgments will not be presented, since these trials did not involve an anchoring manipulation. Comparably, analyses of the between-participants variable block order (two levels: anchoring trials first, nonanchoring trials first) are not presently germane and will not be discussed. The relevant dependent variables for each cell of the design in each experiment are the mean absolute confidence ratings, the mean time to decide absolute confidence, and the mean time to decide relative confidence (i.e., mean time to evaluate the anchor). Signal strength (five levels) served as a within-participants independent variable, and relative confidence level (five levels), absolute confidence level (six levels), and the numeric distance between these two confidence levels also served as within-participants independent variables where appropriate. In each of the following ANOVAs, Huynh-Feldt epsilon adjustment degrees of freedom are used, but the degrees of freedom reported are those defined by the design. An alpha level of .05 was used as an index of significance in all reported analyses.

EXPERIMENT 1

In this experiment, the values used for both the relative and the absolute judgments of confidence are numeric values. Such anchoring conditions are analogous to those within the classic anchoring paradigm (Tversky & Kahneman, 1974) and allow for an initial determination of whether or not confidence judgments are indeed susceptible to classic anchoring effects.

Method

Participants. Six male and 14 female first-year undergraduate students participated in this experiment. One participant failed to complete the task and was replaced.

Apparatus. For this experiment, the seven confidence response buttons were labeled "X," "50," "60," "70," "80," "90," and "100."

Procedure. Participants were told that a rating of 50 was indicative of a *guess*, a rating of 100 indicated *certainty*, and the ratings 60–90 were to be used accordingly. Participants were further instructed to select a confidence rating of X if they were certain that they had made a mistake on the initial signal detection task. The an-

chor values used in this experiment were the midpoints between adjacent values on the confidence scale (i.e., 55, 65, 75, 85, and 95).

Results

Of the 16,000 trials (800 trials \times 20 participants), 1.04% were censored as per the criteria described in the General Method section.

Absolute confidence. There was a significant effect of signal strength on mean absolute confidence level [$F(4,72) = 60.53, p \leq .001, \eta_p^2 = .771$]. It is clear from the plots of Figure 1 that mean absolute confidence increased systematically as a function of signal strength for all trials in which the correct response was *yes* (i.e., where signal strength was greater than 50%).

Importantly, the level of relative confidence did indeed serve as an anchor for participants as they rendered their absolute confidence, with reliably greater absolute confidence being expressed as the level of relative confidence increased [$F(4,72) = 15.28, p \leq .001, \eta_p^2 = .459$]. The relative confidence level \times signal strength interaction was also significant [$F(16,288) = 19.58, p \leq .001, \eta_p^2 = .125$], with the 56% density level being the only strength level for which relative confidence did not appear to act as an anchor for absolute confidence.

Times to render absolute confidence. Only relative confidence levels had a significant effect on the mean time to render absolute confidence judgments [$F(4,72) = 5.57, p \leq .001, \eta_p^2 = .236$]. Participants took reliably more time to make absolute confidence judgments when the anchors were extreme (464 and 463 msec for the 55 and 95 anchor values, respectively) relative to when the anchors were drawn from the middle of the confidence scale (440, 416, and 407 msec for the 65, 75, and 85 anchor values, respectively).

Times to render relative confidence. Mean times to render relative confidence judgments varied systematically with signal strength [$F(4,72) = 20.93, p \leq .001, \eta_p^2 = .538$]. Participants took reliably less time to make a relative confidence decision as signal strength increased.

Interestingly, as demonstrated in Figure 2, a numerical-distance-like effect emerged (Moyer & Landauer, 1967). Participants made relative confidence judgments with reliably greater speed as the numerical difference between the anchor values and the eventual absolute confidence values increased. Because not every participant utilized every absolute confidence value at every level of signal strength for every relative confidence anchor value, it was not possible to analyze this effect using a univariate within-participants ANOVA. As such, a hierarchical linear modeling (HLM) analysis was conducted instead. The HLM model Level 2 clusters were participants, whereas the predictors used at Level 1 of the model were numerical distance (i.e., the absolute value of the difference between the anchor value and the absolute confidence judgment), signal strength, and absolute confidence level. Time to decide relative confidence was the model's dependent variable. Confirming the reliability of the observed distance effect, this analysis revealed that the slope coefficient for the numerical distance predictor was negative (-9.19) and significantly different from 0 [$t(19) = -7.76, p \leq .0001$].

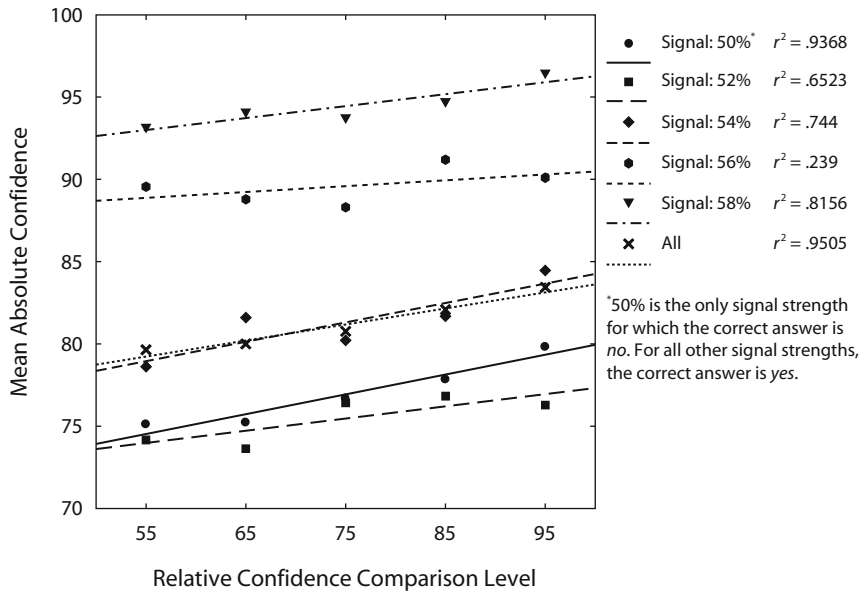


Figure 1. Anchoring effects in Experiment 1: Mean absolute confidence ratings as a function of relative confidence anchor value for each level of signal strength. Separate regression lines for absolute confidence as a function of the relative confidence anchor values are included for each level of signal strength (along with the corresponding r^2 value).

The reliability of this distance effect confirms the semantic comparability of the relative confidence anchor values and the values on the absolute confidence scale (Lammertyn & Fias, 2005). Although there was no reason to assume that such comparability would not exist in this experiment, given that the anchors were the midpoints of the absolute confidence scale values, in the following experiments the presence of such an effect will be an important indicator of semantic comparability.

Discussion

This experiment provided clear evidence that numeric confidence judgments can be influenced by numeric anchors. It also demonstrated that such confidence judgments are not entirely arbitrary. Although they are prone to anchoring effects, they are nevertheless also influenced by decisional difficulty, since stronger signals produced reliably greater confidence judgments. Furthermore, the observed numerical distance effects in the times to make the relative

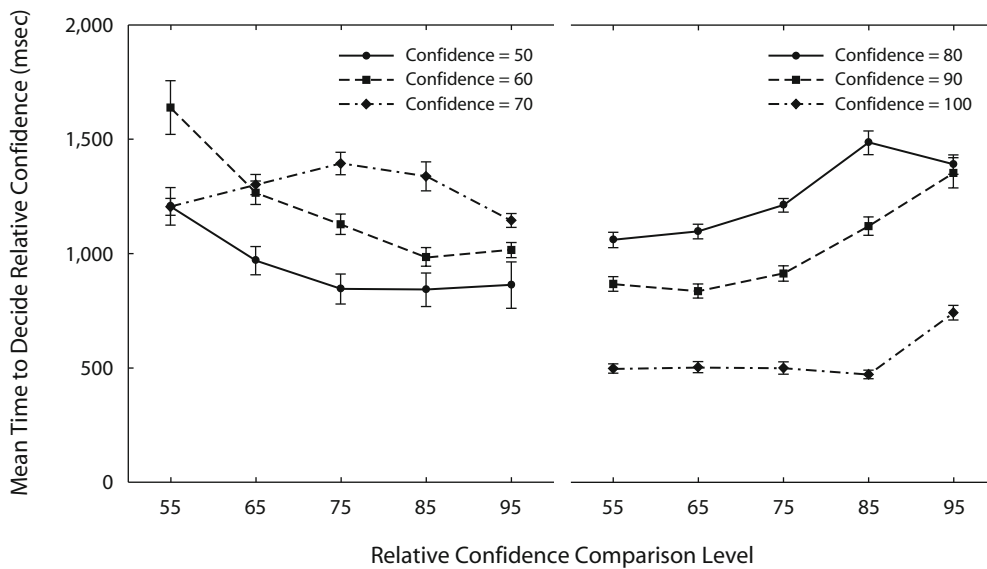


Figure 2. Numerical distance effects in Experiment 1: Time to decide relative confidence as a function of the anchor values for each subsequently rendered absolute confidence level. Error bars represent the standard errors of the means.

confidence judgments indicate that the absolute confidence judgments are indeed based on genuine magnitudes to which proposed numerical anchors can also be compared.

EXPERIMENT 2

The participants in Experiment 1 were asked to use a numeric scale for their absolute confidence judgments, and the anchors used in that experiment were the mid-points between the values of that same scale. If anchoring effects are solely the result of numeric priming, one would not expect to see these effects in a comparable experiment in which both the absolute confidence and relative confidence anchor values are drawn from a nonnumeric scale. On the other hand, if anchoring effects are mainly due to semantic priming, one would still expect to find anchoring effects in this situation, since the values of this scale refer to semantic representations of magnitude. Hence, the purpose of Experiment 2 was to determine whether such a scale could also produce anchoring effects.

Method

Participants. Four male and 16 female first-year undergraduate students participated in this experiment.

Apparatus. For this experiment, six of the confidence response buttons were labeled (in order) as “A,” “C,” “E,” “G,” “I,” and “K.” The seventh, leftmost, button was labeled with a red dot, approximately $\frac{3}{4}$ cm in diameter.

Procedure. The procedure was identical to that used in Experiment 1, except for the type of scale used. Participants were told that a rating of A was indicative of certainty, a rating of K indicated a guess, and the ratings C, E, G, and I were to be used as the participant saw fit, provided that they remembered the rule “letters closer to A indicate greater confidence, letters closer to K indicate greater uncertainty.” (The letter A was chosen to represent certainty because our participants were drawn from a pool of first-year students who, in most cases, associate letters closer to the beginning of the alphabet with academic success.) Participants were further instructed to select the red dot instead of making an absolute confidence judgment if they were certain that they had made a mistake in the initial perceptual-decision task. The relative confidence values that served as the anchors in this experiment were the letters between the ones on the response panel (i.e., B, D, F, H, and J). Participants were told to press the “<” button if they felt that they were less confident than the presented letter and to press the “>” button if they felt that they were more confident than this letter.

Results

Of the 16,000 trials, 3.23% were censored as per the criteria described in the General Method section. The analyses performed directly paralleled those conducted for Experiment 1. In order to perform analyses of mean absolute confidence, though, the six alphabetic values of the scale used by the participants (A, C, E, G, I, and K) were transformed into the comparable numeric scale values used in Experiment 1 (i.e., to 100, 90, 80, 70, 60, and 50, respectively). The relative confidence anchors used in this experiment (B, D, F, H, and J) were likewise transformed (i.e., to 95, 85, 75, 65, and 55, respectively) for analyses of relative confidence judgment times.

Absolute confidence. Mean absolute confidence varied reliably as a function of signal strength [$F(4,72) = 16.59, p \leq .001, \eta_p^2 = .480$]. As was the case in Experi-

ment 1, mean confidence increased along with signal strength for all trials in which the correct response was *yes* (i.e., trials in which signal strength was greater than 50%).

Once again, anchoring effects were observed: Relative confidence levels did indeed serve as anchors for the absolute confidence judgments [$F(4,72) = 6.008, p \leq .002, \eta_p^2 = .250$]. Interestingly, and in spite of the fact that no anchoring effects for the easiest decisions (i.e., the 58% density level) seem to be evident in Figure 3, the relative confidence level \times signal strength interaction was not statistically significant [$F(16,288) = 1.37, p \leq .19$].

Times to render absolute confidence and relative confidence. Both mean times to render absolute confidence judgments [$F(4,72) = 3.09, p \leq .03, \eta_p^2 = .146$] and mean times to render relative confidence judgments [$F(4,72) = 7.44, p \leq .001, \eta_p^2 = .292$] varied systematically with signal strength. Participants took reliably less time to make each type of judgment as signal strength increased.

As in Experiment 1, participants made relative confidence judgments with greater speed as the distance between the anchor values and the eventual absolute confidence values increased (see Figure 4). An HLM analysis analogous to the one performed for Experiment 1 indicated that this distance effect was reliable and revealed that the slope coefficient for the distance predictor was both negative (-16.167) and significantly different from 0 [$t(19) = -4.90, p \leq .001$]. As before, the reliability of this distance effect confirms the semantic comparability of the relative confidence anchor values and the values of the absolute confidence scale used in this experiment.

Discussion

In this experiment, the absolute confidence judgments were influenced by the anchors, even though the values for both were drawn from a nonnumeric scale. As such, the results of this experiment provide the strongest evidence to date—that we are aware of—that such effects can occur independently of the usage of numbers (i.e., numbers are not actually needed to produce anchoring effects, and similarly, it certainly is not the case that the anchoring process needs to be initiated by a numerical prime, as had been suggested by Wong & Kwong, 2000). Furthermore, the observed distance effects in the relative confidence judgment times again suggest that absolute confidence is based on a genuine magnitude with which a proposed anchor can be compared (even when those anchor values are based on purely ordinal-level, semantic representations of magnitude).

EXPERIMENT 3

Because the results of Experiment 2 replicated the findings of Experiment 1 using a nonnumeric scale, it is reasonable to wonder whether anchoring effects can occur when semantically comparable but modally different relative and absolute confidence scales are utilized within each trial (i.e., one nonnumeric and the other numeric). As noted, Chapman and Johnson (1994) found that using different scales for the relative and absolute judgments did not produce any anchoring effects and suggested that the reason for this finding may have been that the scales used

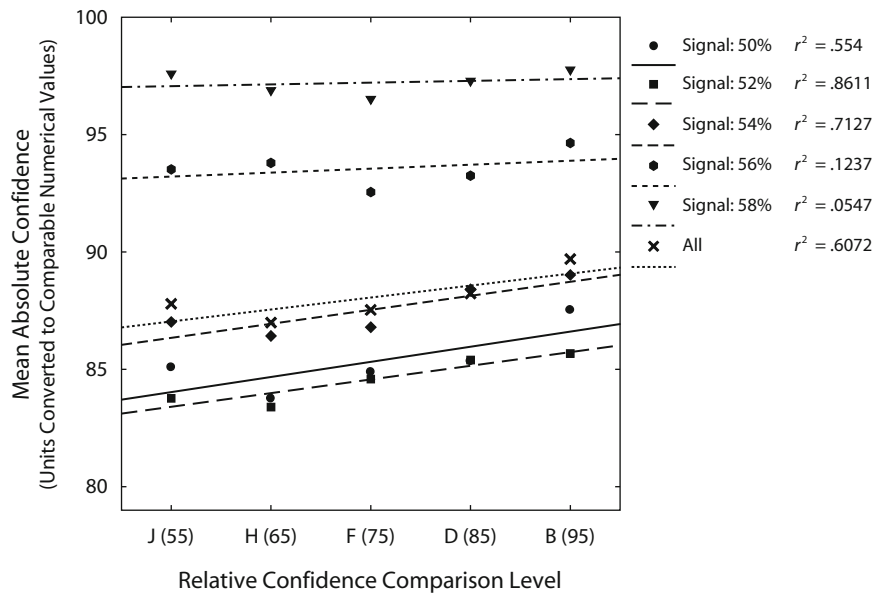


Figure 3. Anchoring effects in Experiment 2: Mean absolute confidence ratings as a function of relative confidence anchor value for each level of signal strength. Separate regression lines for absolute confidence as a function of the relative confidence anchor values are included for each level of signal strength (along with the corresponding r^2 value). Note that the nonnumeric absolute confidence scale used by the participants was converted to the comparable numeric scale from Experiment 1 in order to allow for the generation of mean absolute confidence values. The letters along the x-axis are the nonnumeric anchors used in this experiment. Equivalent numeric scale values are in parentheses.

were not semantically comparable. In Experiment 3, we tested this hypothesis directly.

Method

Participants. Six male and 14 female first-year undergraduate students participated in this experiment. One participant failed to complete the experiment and was replaced.

Apparatus. The seven confidence response buttons on the response panel were labeled as in Experiment 1 (i.e., “X,” “50,” “60,” “70,” “80,” “90,” and “100”).

Procedure. The relative confidence values used in this experiment were *near guess*, *low*, *moderate*, *high*, and *near certain*. In other words, the participants in Experiment 3 used a nonnumeric scale when judging their relative confidence but a numeric scale when judging their absolute confidence. In order to ensure that the participants in this experiment understood that the two scales were semantically congruous, they were explicitly informed that “a confidence rating of 100 indicates certainty, while a confidence rating of 50 indicates a guess.”

Results

Of the 16,000 trials, 3.32% in this experiment were censored. In order to perform analyses of relative confidence judgment times, the anchor values presented to the participants were transformed into the comparable numeric anchor values used in Experiment 1.

Absolute confidence. Once again, absolute confidence increased systematically as a function of signal strength [$F(4,72) = 21.40, p \leq .001, \eta_p^2 = .543$]. The form of this relationship directly parallels that of Experiments 1 and 2.

Importantly, no consistent anchoring effects were observed in Experiment 3 (see Figure 5). Relative confidence levels had no reliable effect on absolute confidence levels. The interaction between relative confidence level and signal strength was significant, however [$F(16,288) = 1.89, p \leq .032, \eta_p^2 = .095$]. Simple effect analyses revealed that relative confidence level had a marginal effect on mean absolute confidence only at the 54%-density signal strength level [$F(4,72) = 2.34, p \leq .068, \eta_p^2 = .115$].

Times to render absolute confidence and relative confidence. Again, both mean times to render absolute confidence judgments [$F(4,72) = 5.33, p \leq .002, \eta_p^2 = .228$] and mean times to render relative confidence judgments [$F(4,72) = 15.93, p \leq .001, \eta_p^2 = .470$] varied systematically with signal strength. The form of these relationships also directly parallels those of Experiments 1 and 2.

Participants were informed a priori of the semantic comparability of the relative and absolute confidence scales (recall that they were explicitly told that 50 was indicative of a guess and that 100 was indicative of certainty), and the presence of distance effects in Experiment 3, which are analogous to those found in the previous two experiments, certainly serves to corroborate the notion that these two scales were indeed semantically comparable. With respect to this point, the distance effects observable in Figure 6 were shown to be reliable via an HLM analysis. This analysis indicated that the slope coefficient for the distance predictor was again negative (-8.37) and significantly different from 0 [$t(19) = -2.93, p \leq .01$]. Although it is

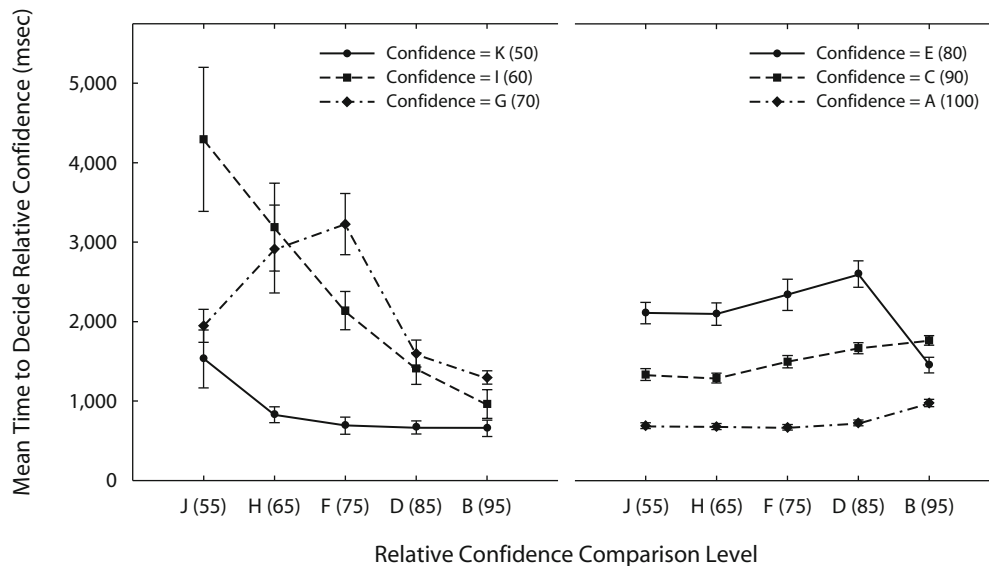


Figure 4. Distance effects in Experiment 2: Time to decide relative confidence as a function of the anchor values for each subsequently rendered absolute confidence level. Error bars represent the standard errors of the means. The letters along the x-axis are the nonnumeric anchors used in this experiment. Equivalent numeric scale values are in parentheses.

not likely the case here (or in either of the two previous experiments, for that matter) that participants actually compared their absolute confidence values with the relative confidence anchor values directly, it is the case that the feelings of confidence invoked on each trial were related to an internal magnitude value that then had comparable effects on both types of judgment (e.g., if this magnitude were relatively large, it would result in both a relatively larger absolute judgment and a faster relative judgment when the anchor value was relatively small).

Discussion

In this experiment, absolute confidence judgments were not influenced by anchors when the values for both were drawn from differing scales, even though the fact that the semantic distance between the absolute judgment and the anchor values affected the relative judgment times indicated that the two scales were indeed semantically comparable. Hence, these findings indicate that semantic comparability was insufficient to cause anchoring effects. As such, they suggest the presence of an important boundary condition regarding the circumstances under which anchoring effects can be expected to be produced by semantic priming—namely, only when the measurement modes of the scales used for the relative and absolute judgments are the same. (Note that in all of the experiments reported previously, both judgments were always numerically based, and semantic comparability was manipulated by varying the semantic congruity between the targets involved with each type of judgment.)

EXPERIMENT 4

To ensure that the findings in Experiment 3 were not simply an artifact of the fact that a nonnumeric scale was

used to evaluate relative confidence and a numeric scale was used to provide absolute confidence, the purpose of Experiment 4 was to replicate those findings with the scales switched. Hence, the participants in this experiment now used a numeric scale when judging their relative confidence and a nonnumeric scale when judging their absolute confidence.

Method

Participants. Five male and 15 female first-year undergraduate students participated in this experiment. One participant failed to complete the experiment and was replaced.

Apparatus. The response panel's confidence response buttons were now labeled as X, GUESS, LOW, MED-LOW, MED-HIGH, HIGH, and CERTAIN.

Procedure. The relative confidence categories used in this experiment were 55, 65, 75, 85, and 95. In other words, participants used a numeric scale when judging relative confidence but a nonnumeric scale when judging absolute confidence. As was the case in Experiment 3, in order to ensure that the participants understood that these two scales were semantically congruous, they were explicitly informed that "a confidence rating of 100 indicates certainty, while a confidence rating of 50 indicates a guess."

Results and Discussion

Of the 16,000 trials, 3.52% were censored. In order to perform analyses of mean absolute confidence, the scale values used by the participants were transformed into the comparable numeric scale values used in Experiment 1.

Absolute confidence. Once again, absolute confidence increased systematically as a function of signal strength [$F(4,72) = 30.71, p \leq .001, \eta_p^2 = .630$]. This relationship is comparable to those described previously.

Once again, no consistent anchoring effects were observed (see Figure 7). The interaction between relative confidence level and signal strength was again significant [$F(16,288) = 2.39, p \leq .01, \eta_p^2 = .118$], and simple effect

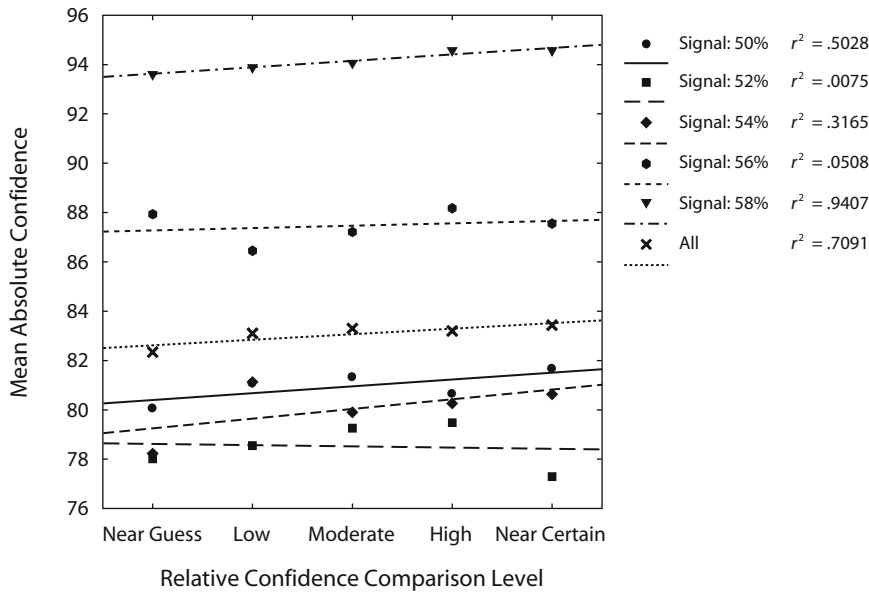


Figure 5. Anchoring effects in Experiment 3: Mean absolute confidence ratings as a function of relative confidence anchor value for each level of signal strength. Separate regression lines for absolute confidence as a function of the relative confidence anchor values are included for each level of signal strength (along with the corresponding r^2 value).

analyses revealed an effect of relative confidence on absolute confidence at the 50%- and 56%-density signal strength levels [$F(4,72) = 6.50, p \leq .001, \eta_p^2 = .265$, and $F(4,72) = 2.88, p \leq .03, \eta_p^2 = .138$, respectively]. An examination of Figure 7, though, reveals that the simple effect of relative confidence on absolute confidence at the 56%-density level cannot be an anchoring effect, because the slope of the best-fitting regression line is negative. Also worth noting in Figure 7 are the low r^2 values (compare these with the r^2 values listed in Figures 1, 3, and 5). Within this experiment, little of

the variability in mean absolute confidence was explained by relative confidence level. As such, there is reason to question the predictive utility of these numeric anchors.

Times to render absolute confidence and relative confidence. Again, both mean times to render absolute confidence judgments [$F(4,72) = 3.82, p \leq .018, \eta_p^2 = .175$] and mean times to render relative confidence judgments [$F(4,72) = 9.33, p \leq .001, \eta_p^2 = .341$] varied systematically with signal strength. These relationships are comparable to those described previously.

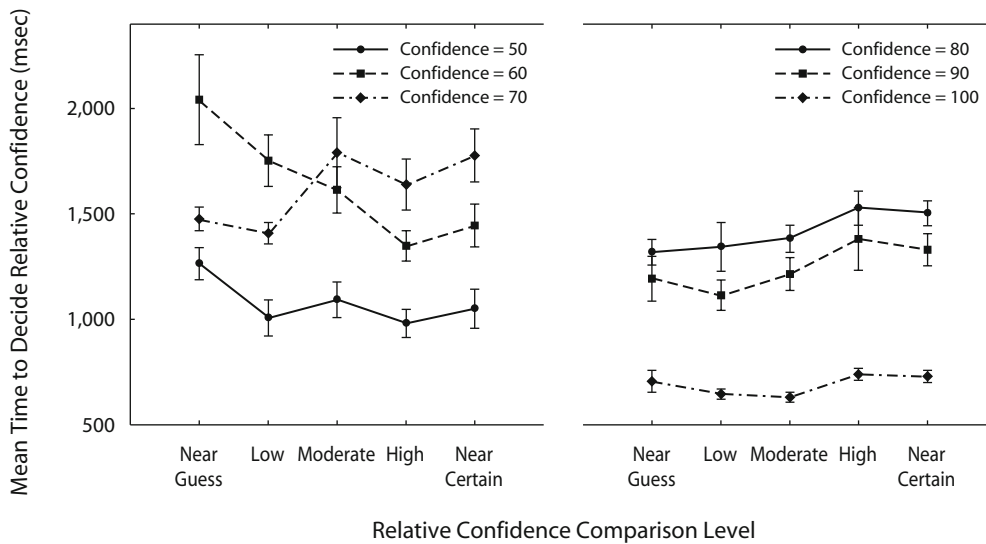


Figure 6. Distance effects in Experiment 3: Time to decide relative confidence as a function of the anchor values for each subsequently rendered absolute confidence level. Error bars represent the standard errors of the means.

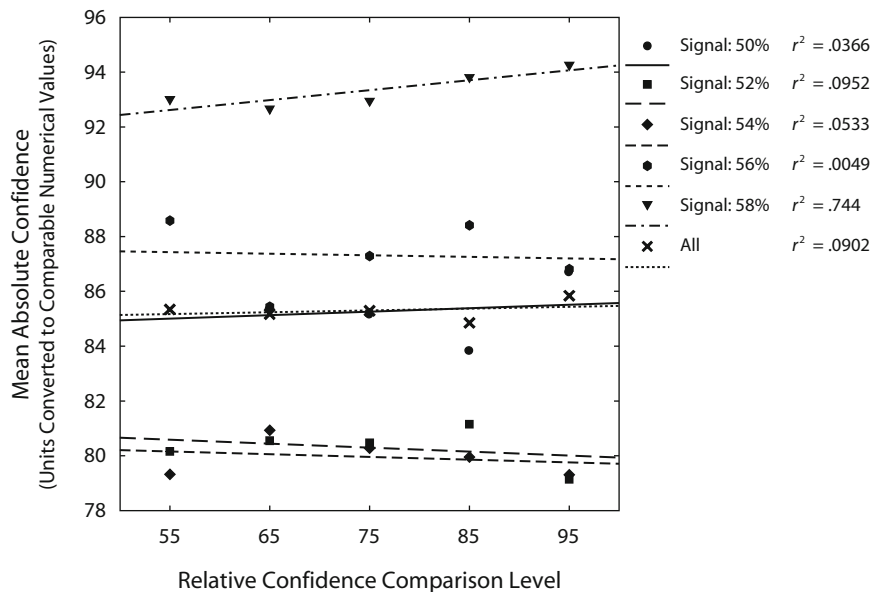


Figure 7. Anchoring effects in Experiment 4: Mean absolute confidence ratings as a function of relative confidence anchor value for each level of signal strength. Separate regression lines for absolute confidence as a function of the relative confidence anchor values are included for each level of signal strength (along with the corresponding r^2 value). Note that the nonnumeric absolute confidence scale used by the participants was converted to the comparable numeric scale from Experiment 1 in order to allow for the generation of mean absolute confidence values.

An HLM analysis indicated that the distance effects observable in Figure 8 were reliable, thus confirming the semantic comparability of the relative and absolute confidence scales. In this analysis, the slope coefficient for the distance predictor was negative (-17.25) and significantly different from 0 [$t(19) = -7.80, p \leq .0001$].

Discussion

In Experiment 4, the finding from Experiment 3 that mere semantic comparability of the absolute and relative judgment scales is insufficient to cause anchoring effects was replicated. Interestingly, this finding occurred even though the scale used for the anchor values was explicitly numeric.

GENERAL DISCUSSION

Experiments 1 and 2 provide clear evidence that confidence judgments can indeed be subject to anchoring effects. With respect to this finding, it is worth mentioning that some other researchers have demonstrated the existence of anchoring effects in metacognitive evaluations. Scheck, Meeter, and Nelson (2004) asked participants to render judgments of learning (JOLs) of paired Swahili–English words and correctly hypothesized that participants would use past performance in similar tasks to judge their own ability to perform the experimental task. As such, these participants were likely generating internal anchors that influenced their JOLs. Although this research provided some support for the notion that metacognitive evaluations can be influenced by anchors, note

that these researchers did not provide participants with explicit anchors, nor did they ask participants to provide any relative (greater than/less than) JOLs. Hence, until now, anchor effects with metacognitive judgments had yet to be demonstrated using the paradigm of a relative judgment followed by an absolute judgment, which has become the hallmark of much of the research on anchoring effects. Nonetheless, our findings do converge nicely with those of Scheck et al. and potentially allow researchers studying metacognitive processes a voice in the debate regarding the causes of anchoring effects.

As such, the fact that quite similar anchoring effects occurred in Experiment 2 for judgments made using alphabetic scale values indicates that the presence of such effects does not actually require the involvement of numerical judgments. Moreover, when the mode of measurement of the relative and absolute confidence scales was varied within an experiment, anchoring effects were not found. This was true both when the anchors were semantic representations of magnitude (Experiment 3) and when the anchors were explicitly numeric (Experiment 4). Along with the findings reported by Chapman and Johnson (1994), these results necessarily place limits on theories of the origin of anchoring effects, because neither the numeric nor the semantic priming view would now seem to provide a sufficient explanation for the anchoring effect. Of primary importance, it seems, is the compatibility of the scales from which the relative and absolute magnitude values are drawn.

The presence of reliable distance effects in the times to make the relative judgments, we believe, does serve to

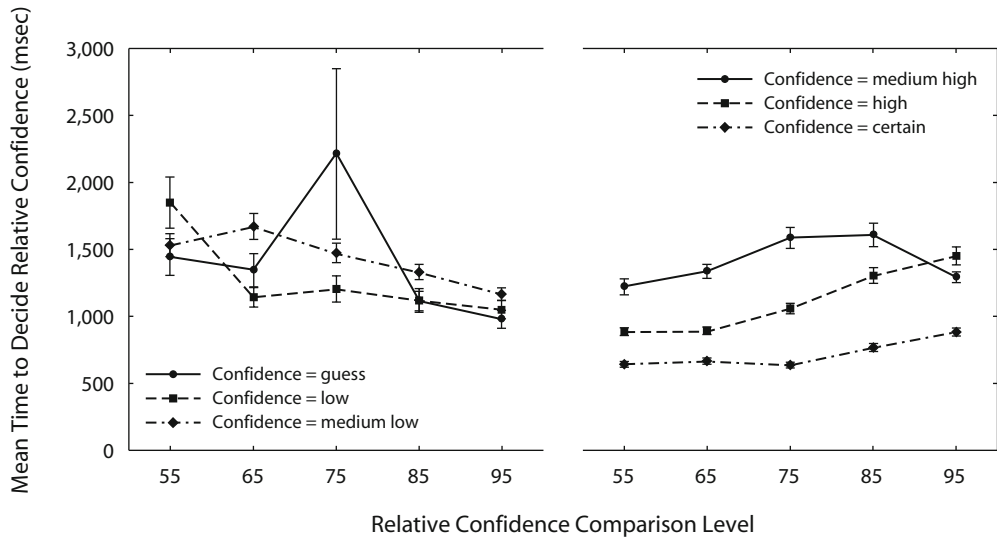


Figure 8. Distance effects in Experiment 4: Time to decide relative confidence as a function of the anchor values for each subsequently rendered absolute confidence level. Error bars represent the standard errors of the means.

confirm the semantic comparability of each of the tested relative confidence scales with the corresponding absolute confidence scales in all four experiments. Although the responses to both types of judgment task must invariably be based on internal magnitude representations of subjective confidence (even for the cases in which the scales used are the same), the fact that such distance effects occur indicates that these internal magnitudes are indeed mapping onto the scales used for each judgment in a similar fashion (i.e., the same internal magnitude that leads to a numeric absolute confidence rating of 60, say, is also quite far away from a semantic anchor value of *near certain*).

Under such circumstances, we also believe that the semantic priming hypothesis, as iterated by Strack and Mussweiler (1997; Mussweiler & Strack, 2001), should clearly have predicted the presence of anchoring effects in Experiments 3 and 4. According to the semantic priming hypothesis, anchoring effects arise because the process of judging the anchor value enhances the subsequent accessibility of confirming anchor-consistent information during the absolute judgment process. Hence, given the demonstrated semantic comparability of the numeric and nonnumeric scales used here, surely the generation of semantic information regarding the degree to which confidence might indeed be consistent with, say, an anchor value of *near certain* that is then rendered highly accessible during the following absolute confidence judgment should serve to affect the value of that judgment, regardless of whether it is made using the same, nonnumeric scale or a comparable numeric one.

This is not to say that the present study discounts the effects of either numeric or semantic priming. Both numeric and semantic anchors were sufficient to cause anchoring effects in the first two experiments. What the results of this study do suggest, however, is that although the means by which anchors exert their effect within any particular

scale domain likely still involve some form of priming, it would seem (for reasons that cannot completely be determined on the basis of the work performed here) that anchor-consistent information that becomes primed with respect to anchors drawn from one type of scale is not necessarily applicable (or transferable) to subsequent absolute judgments made using another type of scale.

Finally, although anchoring effects have typically been explored by having participants make estimates about quantities or extents within the general knowledge domain (for which the exact answer is not likely to be known with any certainty; e.g., what is the population of Toronto?), there is every reason to believe that the anchoring effects observed here are analogous. As noted previously, anchoring effects are pervasive and can become manifest whenever a participant is asked to estimate a quantity and compare that value with a fixed anchor. As such, one would expect to observe anchoring effects whenever the following two conditions are met. First, the target of inquiry must involve the estimation of some quantity, so that the decision maker will not be completely certain of the magnitude being compared with the anchor. Second, it must be possible to position the magnitude of the thing being estimated on some form of linear scale, because an anchoring effect cannot occur if it is impossible to tell whether the anchor is greater or less than that which is being estimated.

We believe that judgments of confidence meet both of these requirements. First, confidence ratings are indeed estimated magnitudes (for which it could also be argued that there never really is any absolute correct answer). Given that they are estimates of subjective experience, they are quite malleable. In fact, participants who are asked to evaluate their certainty of having made a correct choice following each decision in a series of identical decisions nevertheless express quite variable levels of confidence (Baranski & Petrusic, 1994; Henmon, 1911). Second, participants do

seem capable of arranging their rated confidence along a linear scale. Participants in perceptually based decision-making experiments are homogeneously and significantly more confident when making objectively easier decisions (cf. the finding that mean confidence increased with signal strength in all four of the experiments in this study). Beyond any empirically derived data, common experience suggests that it is possible to be more confident in having made a correct decision in one instance and less confident in another (and the adjectives *more* or *less* can be ascribed to the word *confident* without obfuscating its meaning). As such, it seems clear that confidence ratings are indeed estimates that do represent discrete points along a linear continuum and, hence, should be as susceptible to anchoring effects as are standard general-knowledge-based estimates.

AUTHOR NOTE

This work was supported by a Natural Sciences and Engineering Research Council of Canada Graduate Scholarship to S.R.C. and Natural Sciences and Engineering Research Council of Canada Individual Discovery grants to W.M.P. and C.L.-S. We are grateful to Judith Godin for suggesting the use of HLM. Correspondence concerning this article should be sent to S. R. Carroll, W. M. Petrusic, or C. Leth-Steensen, Department of Psychology, Carleton University, Ottawa, ON, K1S 5B6 Canada (e-mail: srcarrol@connect.carleton.ca, bill_petrusic@carleton.ca, or craig_leth_steensen@carleton.ca).

REFERENCES

- BARANSKI, J. V., & PETRUSIC, W. M. (1994). The calibration and resolution of confidence in perceptual judgments. *Perception & Psychophysics*, *55*, 412-428.
- CHAPMAN, G. B., & JOHNSON, E. J. (1994). The limits of anchoring. *Journal of Behavioral Decision Making*, *7*, 223-242. doi:10.1207/s15516709cog2901_5
- HENMON, V. A. C. (1911). The relation of the time of a judgment to its accuracy. *Psychological Review*, *18*, 186-201. doi:10.1037/h0074579
- JACOWITZ, K. E., & KAHNEMAN, D. (1995). Measures of anchoring in estimation tasks. *Personality & Social Psychology Bulletin*, *21*, 1161-1166. doi:10.1177/01461672952111004
- LAMMERTYN, J., & FIAS, W. (2005). Negative priming with numbers: No evidence for a semantic locus. *Quarterly Journal of Experimental Psychology*, *58A*, 1153-1172. doi:10.1080/02724980443000520
- MOYER, R. S., & LANDAUER, T. K. (1967). Time required for judgments of numerical inequality. *Nature*, *215*, 1519-1520.
- MUSSWEILER, T., & STRACK, F. (2001). The semantics of anchoring. *Organizational Behavior & Human Decision Processes*, *86*, 234-255. doi:10.1006/obhd.2001.2954
- SCHECK, P., MEETER, M., & NELSON, T. O. (2004). Anchoring effects in the absolute accuracy of immediate versus delayed judgments of learning. *Journal of Memory & Language*, *51*, 71-79. doi:10.1016/j.jml.2004.03.004
- STRACK, F., & MUSSWEILER, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality & Social Psychology*, *73*, 437-446. doi:10.1037/0022-3514.73.3.437
- TVERSKY, A., & KAHNEMAN, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124-1131. doi:10.1126/science.185.4157.1124
- WONG, K. F. E., & KWONG, J. Y. Y. (2000). Is 7300 m equal to 7.3 km? Same semantics but different anchoring effects. *Organizational Behavior & Human Decision Processes*, *82*, 314-333. doi:10.1006/obhd.2000.2900

(Manuscript received February 23, 2008;
revision accepted for publication August 8, 2008.)