

Duration Sensitivity Depends on Stimulus Familiarity

Carey K. Morewedge
Carnegie Mellon University

Karim S. Kassam
Harvard University

Christopher K. Hsee and Eugene M. Caruso
University of Chicago

When people are asked to assess or compare the value of experienced or hypothetical events, one of the most intriguing observations is their apparent insensitivity to event duration. The authors propose that duration insensitivity occurs when stimuli are evaluated in isolation because they typically lack comparison information. People should be able to evaluate the duration of stimuli in isolation, however, when stimuli are familiar and evoke comparison information. The results of 3 experiments support the hypothesis. Participants were insensitive to the duration of hypothetical (Experiment 1) and real (Experiment 2) unfamiliar experiences but sensitive to the duration of familiar experiences. In Experiment 3, participants were insensitive to the duration of an unfamiliar noise when it was unlabeled but sensitive to its duration when it was given a familiar label (i.e., a telephone ring). Rather than being a unique phenomenon, duration neglect (and perhaps other forms of scope insensitivity) appears to be a particular case of insensitivity to unfamiliar attributes.

Keywords: duration neglect, magnitude estimation, evaluability, joint and single evaluation, judgment and decision making

Intuition suggests that more of a good thing is invariably better and more of a bad thing is invariably worse. But when stimuli are judged in isolation, people are often insensitive to scope variables such as magnitude and duration. People are willing to pay an equal amount of money to prevent 2,000, 20,000, or 200,000 birds from drowning in oil (Desvovges et al., 1993) and sometimes prefer longer rather than shorter ice baths, colonoscopies, and unpleasant films (for reviews, see Ariely, Kahneman, & Loewenstein, 2000; Fredrickson, 2000; Roy, Christenfeld, & McKenzie, 2005). Yet, there are limits to their insensitivity to magnitude and duration. People are sensitive to small differences in the size of requests for charity donations (Gourville, 1998), report greater happiness when receiving larger payments (Morewedge, Gilbert, Keysar, Berkovits, & Wilson, 2007), and report being less satisfied with their lives when they spend more of their time in traffic (Stutzer & Frey, 2004). We propose that when a stimulus is judged in isolation,

sensitivity to scope variables (i.e., their evaluability) depends on the cognitive accessibility of similar stimuli, which in turn depends on stimulus familiarity. In this article, we test this hypothesis within the domain of duration.

Scope Insensitivity

Scope insensitivity is more likely when stimuli are judged in isolation than when they are judged in comparison to other stimuli (for a review, see Hsee, Loewenstein, Blount, & Bazerman, 1999). In one demonstration, research participants were asked how much they were willing to pay for different-sized servings of ice cream. One serving consisted of a 5-oz cup overfilled with 7 oz of ice cream. The other serving consisted of a 10-oz cup underfilled with 8 oz of ice cream. Although people generally prefer more ice cream to less, participants who only saw a single cup of ice cream were willing to pay more for the smaller (overfilled) than the larger (underfilled) serving. This preference reversal presumably occurred because in isolation it is difficult to evaluate magnitude (i.e., the amount of ice cream in a serving) but easy to evaluate whether a serving is overfilled or underfilled (Hsee, 1998).

People exhibit sensitivity to scope in isolation when they possess explicit distributional information about the scope variable such as its range, average, or frequency of specific values or when they possess a modulus to incorporate in their judgments (Hsee et al., 1999; Parducci, 1965, 1995; Stevens, 1975). When one possesses no distributional information about an attribute, it is impossible to evaluate a single value because one has no standard of comparison. With a single point of comparison, an attribute can be evaluated as better or worse, but one will be unable to assess the extent of its superiority or inferiority. If the best and worst values of the distribution of an attribute are known, even small differences

Carey K. Morewedge, Department of Social and Decision Sciences, Carnegie Mellon University; Karim S. Kassam, Department of Psychology, Harvard University; Christopher K. Hsee and Eugene M. Caruso, Booth School Of Business, University Of Chicago.

We gratefully acknowledge the support of Carnegie Mellon University, the Booth School of Business at the University of Chicago, and the John Templeton Foundation. We thank Brittani Baxter, Erica Cohen, Alex Davis, Sarah Kasuga, Amit Kumar, Elizabeth Martin, Lisa Sunner, Sara Rabinovitch, and Alicia Warlick for help with the execution of these experiments and Dan Ariely, Daniel Kahneman, and Daniel T. Gilbert for helpful suggestions.

Correspondence concerning this article should be addressed to Carey K. Morewedge, Department of Social and Decision Sciences, Carnegie Mellon University, 208 Porter Hall, Pittsburgh, PA 15213. E-mail: morewedge@cmu.edu

between values of that attribute can be distinguished. When asked to play the role of a college admissions officer, for example, people given no distributional information about admissions test scores judge candidates with higher and lower scores similarly (Hsee et al., 1999). Given the average test score, people judge candidates with above-average scores to be more desirable than candidates with below-average scores, but they do not distinguish between scores that differ slightly or greatly from the average score. Given the best and worst possible scores, however, people reliably rate the desirability of candidates in accordance with their position in the distribution of candidates. Similarly, when people are initially provided with a standard experience (i.e., an aversive sound) and asked to choose between repeating either that standard experience or novel aversive sounds to which they are subsequently exposed, their choices exhibit duration sensitivity (Ariely & Loewenstein, 2000).¹ Even when people are making comparative judgments, the more explicit distributional information they possess, the greater their scope sensitivity.

Many evaluations are made in isolation without being explicitly compared with a range or standard. There may be a single decent route from home to work, and one may not be able to remember the largest and smallest amount of time one has spent traveling it. People may wonder whether it has been too long since they last spoke to an old friend, consider whether the amount of time spent waiting in a new doctor's office is unusual, have to estimate how much food to order for a dinner party, or decide whether a television is too large for their living room. What determines whether scope variables are evaluable in isolated judgments when their distributions must be inferred?

Evidence from investigations of psychophysical judgment may help to provide an answer. As people convert stimulus intensities into magnitudes according to a lawful form (Fechner, 1860/1966; Stevens, 1975; Weber, 1846/1965), one might expect isolated scope judgments to be relatively consistent between individuals. Magnitude estimations, however, are made by converting physical magnitudes into mental representations of the stimulus according to ratios between stimuli rather than according to absolute differences between stimuli. So the particular stimuli with which a given stimulus is compared can profoundly influence estimations of its magnitude.

Judgments of the same stimulus vary in a systematic fashion according to the range and frequency of stimuli that one has recently experienced (Parducci, 1965). In other words, the relationship between objective magnitudes and magnitude judgments depends on the context in which stimuli are judged—other stimuli that were recently evaluated (Laming, 1997). This suggests that judgments of the scope of a stimulus will exhibit reliability across judges when their judgments are made in the same context, whether that context consists of a single modulus or set of identical stimuli with which the stimulus may be compared.

Not all contexts instill reliability in judgment. Contexts only instill reliability when they consist of stimuli that are similar to the target (Gourville, 1998; Kahneman & Miller, 1986). Stewart and Chater (2003), for example, presented participants with a sinusoidal tone in one ear and a white noise hiss tone of equal loudness in the other and asked them to judge which tone was louder. When the tones were preceded by a similar pair in which the sinusoidal tone was quieter and the white noise hiss was louder, the new sinusoidal tone was judged to be louder (even though both tones

were of equal loudness). Participants thus did not compare the loudness of each new tone with the loudness of both the similar and the different previous tones that they heard—they only compared the loudness of each new tone with the loudness of the similar previous tone that they heard. This suggests that people may only be able to evaluate the scope of a stimulus by comparing it with other similar stimuli.

If the scope of a stimulus is only judged by comparison with other similar stimuli, its scope should be evaluable in isolation only when one is familiar with similar stimuli (so they can be retrieved from memory) and representations of those similar stimuli are cognitively accessible at the time of judgment. The perceived duration of a person's daily commute may thus be evaluated by comparison with his or her most recent, arduous, and/or painless commutes on that particular route, with the accessibility of those similar trips making judgments of the duration of the commute possible and consistent with others' isolated judgments of its duration. In contrast, the perceived duration of a person's trip to a novel location is unlikely to evoke and be compared with a consistent set of other similar trips, making judgments of its duration difficult and inconsistent with others' isolated judgments of its duration.

Duration Insensitivity/Neglect

Duration is a particularly interesting scope variable, because people often appear to neglect it entirely when evaluating experiences. People exhibit surprising insensitivity to the duration of unfamiliar events such as novel sounds and plotless videos (Fredrickson & Kahneman, 1993; Schreiber & Kahneman, 2000), heat probes (Ariely, 1998), cold pressor tasks (Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993), and invasive medical procedures (Redelmeier & Kahneman, 1996), even when judging those stimuli in direct comparison. When experimenters control for the peak and final moments of the pleasant and unpleasantness of films, for example, participants' global evaluations of aversive and pleasant films exhibit insensitivity to duration, even when the durations of shorter and longer clips vary by a factor of three (Fredrickson & Kahneman, 1993). Global ratings of the discomfort felt during colonoscopies and lithotripsy are similarly insensitive to the duration of those procedures: No significant correlations have been found between the length of the procedures and retrospective evaluations (Redelmeier & Kahneman, 1996). Insensitivity to duration is not limited to measures using self-reports—it is also found with behavioral measures. When asked which experience they would prefer to repeat, participants who placed their hand in a 14 °C bath of ice water for 60 s and also in a bath of ice water for 90 s that was 14 °C for the first 60 s and then 15 °C for the next 30 s chose to repeat the latter (longer and more painful) experience (Kahneman et al., 1993).

¹ In contrast, people evaluating or choosing between experiences in several of Kahneman and colleagues' experiments exhibited duration insensitivity. It is possible they exhibited duration insensitivity because they were not provided with a standard or normal distribution of experiences to compare with their experiences (Fredrickson & Kahneman, 1993; Kahneman et al., 1993; Redelmeier & Kahneman, 1996; Schreiber & Kahneman, 2000).

Even when people are sensitive to duration, they are markedly less sensitive to it than to other scope variables such as the intensity of experiences (Ariely & Loewenstein, 2000). Furthermore, the perceived duration of experiences is extremely malleable. For example, it is moderated by contextual factors such as the number of meaningful events that occur during the experience and how the experience is partitioned (Ariely & Zauberman, 2000, 2003; Zakay & Block, 1997).

Although the dominant discounted utility model suggests that experiences should be evaluated by their duration-weighted intensities (Ariely & Loewenstein, 2000; Kahneman, Wakker, & Sarin, 1997), global evaluations of extended experiences appear to neglect their duration and to be best predicted by the intensity of their most intense and final moments. Such results have led some researchers to suggest that people exhibit *duration neglect*, or insensitivity to the duration of pleasant and unpleasant experiences (e.g., Fredrickson, 2000; Fredrickson & Kahneman, 1993; Kahneman et al., 1993; Redelmeier & Kahneman, 1996; Schreiber & Kahneman, 2000; cf. Ariely & Loewenstein, 2000). Duration neglect is purported to be an attentional phenomenon, whereby “what comes most readily to mind in evaluating episodes are the salient moments of those episodes and the affect associated with those moments” (Fredrickson & Kahneman, 1993, p. 54). Indeed, many studies in this line of research show that the peak and end intensities of experiences are their most salient and memorable features (Diener, Wirtz, & Oishi, 2002; Fredrickson, 2000; Fredrickson & Kahneman, 1993; Kahneman et al., 1993; Morewedge, Gilbert, & Wilson, 2005; Schreiber & Kahneman, 2000; Wirtz, Kruger, Scollon, & Diener, 2003). Because duration is typically assumed to be less salient than other factors of an extended experience, it is presumed to affect global judgments of those experiences only when attention is explicitly drawn to it (Ariely, Kahneman, & Loewenstein, 2000).

It is possible, however, that the tendency to neglect duration in previous experiments was largely due to the unfamiliarity of the particular experiences rather than to a more general tendency to neglect duration when evaluating all extended experiences. Duration sensitivity was exhibited in several experiments using factorial designs (Ariely, 1998; Ariely & Loewenstein, 2000; Kahneman, Ritov, & Schkade, 1999; Schreiber & Kahneman, 2000), a finding consistent with the notion that scope sensitivity depends on the cognitive accessibility of a consistent set of stimuli with which the target stimulus can be compared. People readily recall statistical information about the properties of sets of stimuli to which they have been exposed such as their mean and range (Ariely, 2001). If a stimulus (e.g., *brrrrrring*) consistently evokes a particular set of stimuli (e.g., a telephone’s ring), judges should be able to evaluate the duration of that experience, as consistent standards would be cognitively accessible at the time of judgment (e.g., the mean length of a telephone ring). In these experiments with factorial designs, participants could have rated enough stimuli to extract distribution information for the set of stimuli, creating a setting in which the duration of any particular stimulus could have been compared with the mean stimulus duration or range of durations, which would allow them to reliably evaluate the duration of those stimuli.

The Present Research

In three experiments, we examined whether sensitivity to the duration of a stimulus in isolated judgment depends on the cognitive accessibility of a consistent set of stimuli with which it may be compared. We chose to examine duration rather than other scope variables because it provides a particularly conservative test of our hypotheses. We predicted that isolated ratings of familiar experiences would exhibit sensitivity to duration because familiar experiences evoke a consistent set of stimuli. We predicted that participants’ isolated ratings of unfamiliar experiences would exhibit insensitivity to duration unless a consistent set of stimuli with which they may be compared is evoked or provided at the time of judgment. In short, we suggest that the duration of unfamiliar experiences is neglected because they do not evoke consistent standards with which they may be compared rather than because duration is not a salient feature of extended experiences.

Our first two experiments tested whether evaluations would exhibit sensitivity to the duration of familiar but not unfamiliar extended experiences. In Experiment 1, we compared the sensitivity of commuters to the duration of a hypothetical commute along a familiar or an unfamiliar route. In Experiment 2, we compared sensitivity to the duration of unpleasant familiar and unfamiliar sounds of equal length in retrospective evaluations. In both cases, we expected participants to be sensitive to the duration of familiar experiences and be insensitive to the duration of unfamiliar experiences. The third experiment tested whether participants would be sensitive to the duration of a novel stimulus if a category of similar familiar stimuli was made cognitively accessible at the time of judgment. Evaluations of all stimuli were made in isolation without explicitly drawing attention to their duration. The attentional account of duration neglect would thus predict that participants should exhibit duration neglect in all evaluations. We predicted that participants would be insensitive to the duration of unfamiliar stimuli but would be sensitive to the duration of familiar stimuli and unfamiliar stimuli when they were provided with a familiar category with which those stimuli could be compared.

Experiment 1

In a between-subjects design, Experiment 1 tested whether sensitivity to the scope of a hypothetical experience depends on familiarity with that experience. There are two shuttles on the Harvard College campus that transport students from their dormitories to the center of campus. One leaves from a dormitory south of campus; the other leaves from a dormitory north of campus. Residents of the north campus (i.e., the Quad) frequently use the shuttle as a mode of transport, whereas residents of the south campus (i.e., the River houses) do not frequently use the shuttle as a mode of transport.² Students are randomly assigned to dormito-

² We suspect Quad residents more frequently ride the campus shuttle because the walking distance between their dormitory and the center of campus is greater (1.1 miles) than the walking distance between the River dormitories and the center of campus (0.2 miles). The duration of the two shuttle rides may be similar because the shuttle takes a more direct route from the Quad houses to the campus center than it does from the River houses to the campus center; the River houses shuttle also passes through areas of heavy traffic (e.g., Harvard Square).

ries, thus their asymmetric use of the shuttles serves as a natural manipulation of stimulus familiarity.

As the duration of the hypothetical trips received equal mention in all of the vignettes that participants evaluated, an attentional account would predict that participants would exhibit equal sensitivity to the durations of commutes that were familiar and unfamiliar. In contrast, we expected duration to influence evaluations of frequently experienced trips and not to influence evaluations of infrequently experienced trips because only frequently experienced trips would evoke consistent comparison standards. We predicted that Quad residents would be sensitive to the duration of a trip on the Quad shuttle route but not on the River house shuttle route, and we predicted that River house residents would not be sensitive to the duration of a trip on either route.

Method

Participants. Two hundred seven Harvard University undergraduates (148 women, $M_{\text{age}} = 20.1$ years, $SD = 1.4$) who responded to an e-mail to their dormitory electronic mailing list completed an online survey in exchange for the chance to win \$100.

Procedure. Each participant read one vignette about a trip on one of two campus shuttles from a residential area at the north end of campus (i.e., the Quad) or the south end of campus (i.e., Mather house) to the central shuttle stop. Both starting points are equidistant from the central shuttle stop (i.e., Memorial Hall) and have an average duration of 10 min. The vignettes described a shuttle trip that took 12 or 17 min:

Imagine one rainy day you took the Harvard Shuttle from Mather House [the Quad] to Memorial Hall because you lost your umbrella. You had to stand because the shuttle was crowded, and the person next to you bumped into you, poking you with their wet umbrella. To top it off, due to traffic and the number of people trying to get on the Shuttle, the ride took 12 [17] minutes. However, the person who bumped you did apologize for their clumsiness.

At the bottom of that Web page, participants reported how pleasant the experience would be on an 11-point Likert-type scale with the endpoints *very unpleasant* (1) and *very pleasant* (11). On a subsequent Web page, participants reported how frequently they took that shuttle—*never* (1), *at least once a semester* (2), *at least once a month* (3), *at least once a week* (4), *at least once a day* (5)—and the dorm in which they resided. Participants were randomly assigned to the shuttle route and duration conditions.

Results

Participants were split into two residential groups: those who resided in a River house (i.e., the south campus; $n = 141$) and those resided in a house in the Quad (i.e., the north campus; $n = 66$). This distinction is widely recognized on the Harvard campus (Dunn, Wilson, & Gilbert, 2003).

Manipulation check. We submitted participants' usage frequency to a 2 (residence: River, Quad) \times 2 (shuttle route: River, Quad) between-subjects analysis of variance (ANOVA). That analysis revealed that residents of Quad houses took the shuttle more often than did residents of River houses, $F(1, 199) = 9.55$, $p < .01$, $\eta_p^2 = .05$, and the Quad shuttle was more

frequently used, $F(1, 199) = 43.38$, $p < .001$, $\eta_p^2 = .18$. These main effects were qualified by a Residence \times Shuttle Route interaction, $F(1, 199) = 54.61$, $p < .001$, $\eta_p^2 = .22$. As predicted, planned contrasts revealed that residents of Quad houses took the Quad shuttle more often than the River shuttle, $F(1, 199) = 73.74$, $p < .001$, whereas residents of River houses took neither the Quad nor the River shuttle more frequently, $F < 1$. In short, residents of the Quad frequently took the Quad shuttle route and rarely took the River shuttle route, whereas residents of the River houses rarely took either shuttle route (see Figure 1).

Evaluations of the experience. On the basis of frequency of shuttle use, we predicted that evaluations made by residents of the Quad would exhibit sensitivity to the duration of a trip on the Quad shuttle but insensitivity to the duration of a trip on the River shuttle and that evaluations made by residents of the River houses would exhibit insensitivity to the duration of a trip on either route. This is indeed what we found: When assessments of the vignettes' pleasantness were submitted to a 2 (residence: River, Quad) \times 2 (shuttle route: River, Quad) \times 2 (duration: 12 min, 17 min) between-subjects ANOVA, it yielded a significant main effect of shuttle route on pleasantness, $F(1, 199) = 6.26$, $p = .01$, $\eta_p^2 = .03$, which was qualified by the predicted three-way Residence \times Shuttle Route \times Duration interaction, $F(1, 199) = 5.20$, $p = .02$, $\eta_p^2 = .03$.³ Planned contrasts revealed that Quad residents considered a 12-min trip to be more pleasant than a 17-min trip on the Quad shuttle, $F(1, 199) = 4.72$, $p = .03$, but showed no reliable differences in their evaluations of 12- and 17-min trips on the River shuttle, $F(1, 199) = 1.89$, $p = .17$. River house residents showed no reliable differences in the evaluation of 12- or 17-min trips on both the Quad and the River shuttles, $F_s < 1$ (see Figure 1).

Discussion

Although duration was made equally salient in vignettes describing familiar and unfamiliar experiences, duration sensitivity systematically varied according to their familiarity. Consistent with previous research suggesting an attentional basis for duration neglect (Varey & Kahneman, 1992), evaluations of unfamiliar experiences showed no reliable differences. Consistent with our hypothesis, evaluations of familiar experiences were sensitive to their duration. A longer trip appeared less pleasant than a shorter trip to students who traveled that route frequently, but longer and shorter trips appeared similarly unpleasant to students who traveled that route infrequently (or never). This suggests that the familiarity of an experience—rather than the salience of its duration—may determine whether prospective evaluations are sensitive to its duration.

Unusual stimuli and experiences are particularly likely to be both noticed and remembered (Morewedge et al., 2005; Wirtz et al., 2003; Wolfe & Horowitz, 2004). It is possible that the particularly long duration of the longer familiar experiences (i.e., a 17-min commute that normally lasts 10 min) made its extended

³ Levine's test for the equality of variances did not find significant differences in error variance for the dependent variable across groups, $F < 1$.

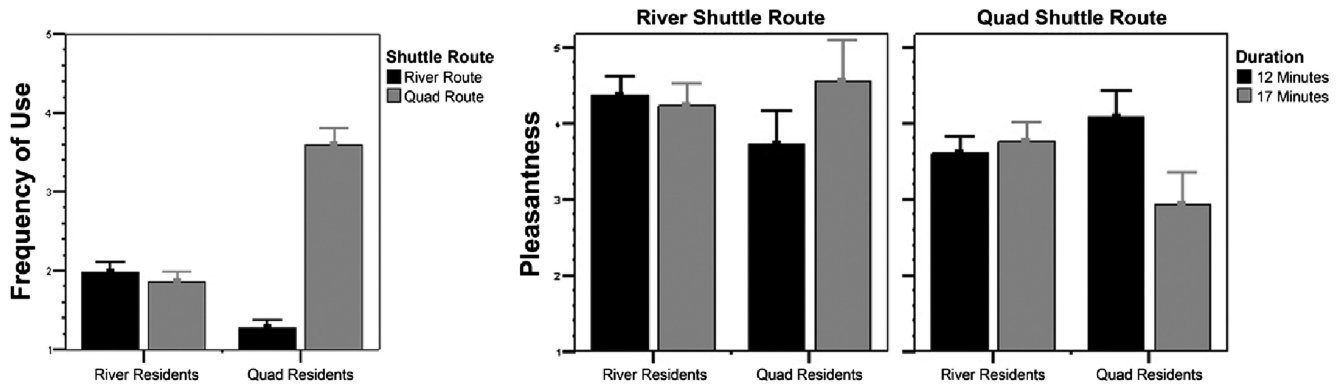


Figure 1. Residents of Harvard Quad houses were familiar with the Quad shuttle route, whereas residents of Harvard River houses were familiar with neither shuttle route (left). When evaluating the pleasantness of a prospective trip in isolation, Quad residents were thus sensitive to the duration of a trip on the Quad shuttle route but not the River shuttle route, whereas River residents were insensitive to the duration of a trip on either shuttle route (right). Bars represent +1 standard error.

duration salient, which in turn led participants to attend to the duration of that experience and evaluate the experience most negatively. Of course, this attentional explanation of their duration sensitivity would require the duration of the typical familiar commute to be cognitively accessible for participants to have noticed the peculiar length of the longer commute. To rule out this possible alternative interpretation of the results, we ensured that no explicit mention of the duration of experiences was made in subsequent experiments.

Experiment 2

Experiment 2 expanded our investigation in three important ways. First, participants were exposed to and evaluated real rather than hypothetical experiences. Second, participants made retrospective evaluations of experiences. Third, no explicit mention of the duration of experiences was made—participants simply heard a longer or shorter audio track consisting of a sound that was unfamiliar (i.e., a synthesizer bleep) or familiar (i.e., a rotary telephone ring). Immediately afterward, participants evaluated the pleasantness of the experience. We predicted that retrospective evaluations of the unfamiliar stimuli would demonstrate duration neglect, whereas retrospective evaluations of the familiar stimuli would demonstrate duration sensitivity.

Method

Participants. One hundred forty-seven pedestrians on the University of Chicago campus (72 women; $M_{\text{age}} = 22.7$ years, $SD = 1.2$) listened to an audio track and completed a survey in exchange for candy. Participants were recruited in the lobby of a student center and were then led to a quiet room in which they listened to the sound and completed the survey.

Materials and stimuli. Participants listened to audio tracks on a SanDisk Sansa 2-GB MP3 player over headphones that did not entirely cover the ear. All audio tracks were played at the same volume. The unfamiliar sound was synthesized. It lasted 2.5 s with cyclic, varying amplitude and a 0.5 s pause between repeats. The familiar sound consisted of a looped rotary telephone ring. It held

at maintaining maximum intensity for 1.7 s and then gradually trailed off over the next 1.3 s, with approximately 0.04 s in between the rings. Both sounds were thus built in 3-s units. The short duration sound included two units and the long duration sound included seven units. Audio tracks were thus 6 s or 21 s in duration.

Pretest. In a between-subjects design, 24 Harvard undergraduates (14 women, $M_{\text{age}} = 21.7$ years, $SD = 3.4$) listened to the familiar or unfamiliar audio track for 6 s on an iRiver 1-GB portable MP3 player and then rated its familiarity and pleasantness on 11-point Likert-type scales marked with endpoints such as *it was very unpleasant* (−5) and *it was very pleasant* (5). The pretest was conducted in a dormitory dining hall during a period set aside for quiet studying. Participants considered the familiar sound to be more familiar than the unfamiliar sound ($M_{\text{familiar}} = 3.08$, $SD = 1.50$; $M_{\text{unfamiliar}} = 0.75$, $SD = 3.02$), $F(1, 22) = 5.74$, $p = .03$, $\eta_p^2 = .21$, but considered both sounds to be similarly pleasant ($M_{\text{familiar}} = 0.08$, $SD = 1.88$; $M_{\text{unfamiliar}} = 1.17$, $SD = 2.94$), $F(1, 22) = 1.15$, $p = .30$.

Procedure. In a between-subjects design, each participant listened to a single familiar or unfamiliar audio track with a short (6 s) or long (21 s) duration. After listening to the audio track, they rated how pleasant or unpleasant that experience was, how familiar or unfamiliar that experience was, and how impatient they felt during the experience on 11-point Likert-type scales marked with endpoints such as *it was very unpleasant* (−5) and *it was very pleasant* (5). Participants were randomly assigned to conditions.

Results

Manipulation check. A 2 (sound: familiar, unfamiliar) \times 2 (duration: short, long) between-subjects ANOVA on participants' reports of familiarity revealed the predicted main effect of sound, $F(1, 143) = 36.42$, $p < .001$, $\eta_p^2 = .20$; there was no main effect of duration and no interaction, $F_s < 1$. Participants who listened to the telephone ring considered the sound they heard to be more familiar ($M = 1.83$, $SD = 3.0$) than did participants who listened to the synthesizer bleep ($M = -0.87$, $SD = 2.4$).

Evaluations of the experience. Evaluations of the pleasantness of the experiences were submitted to a 2 (sound) \times 2 (duration) ANOVA, which yielded the predicted interaction, $F(1, 147) = 4.64, p = .03, \eta_p^2 = .03$, but no main effect of sound or duration, $F < 1$ and $F(1, 147) = 1.35, p = .25$, respectively. Planned contrasts revealed that participants who listened to the longer familiar audio track considered it less pleasant than did participants who listened to the shorter familiar audio track ($M_{\text{longer}} = -0.85, SD = 2.15; M_{\text{shorter}} = 0.45, SD = 1.78$), $F(1, 147) = 5.38, p = .02$, whereas participants who listened to the longer and shorter unfamiliar audio tracks showed no reliable differences in their evaluations ($M_{\text{longer}} = -0.02, SD = 2.78; M_{\text{shorter}} = -0.41, SD = 2.54$), $F < 1$ (see Figure 2).

A 2 (sound) \times 2 (duration) ANOVA on the extent to which participants reported feeling impatient during the experience yielded a main effect of duration, $F(1, 147) = 8.21, p = .005, \eta_p^2 = .05$, which was qualified by the predicted interaction, $F(1, 147) = 6.09, p < .02, \eta_p^2 = .04$. Planned contrasts revealed that participants who listened to the longer familiar audio track reported feeling more impatient during the experience than did participants who listened to the shorter familiar audio track ($M_{\text{longer}} = 0.91, SD = 2.74; M_{\text{shorter}} = -1.63, SD = 2.67$), $F(1, 147) = 13.86, p < .001$, whereas participants who listened to the longer and shorter unfamiliar audio tracks showed no reliable differences in their evaluations ($M_{\text{longer}} = -0.93, SD = 3.14; M_{\text{shorter}} = -1.12, SD = 2.88$), $F < 1$ (see Figure 2).

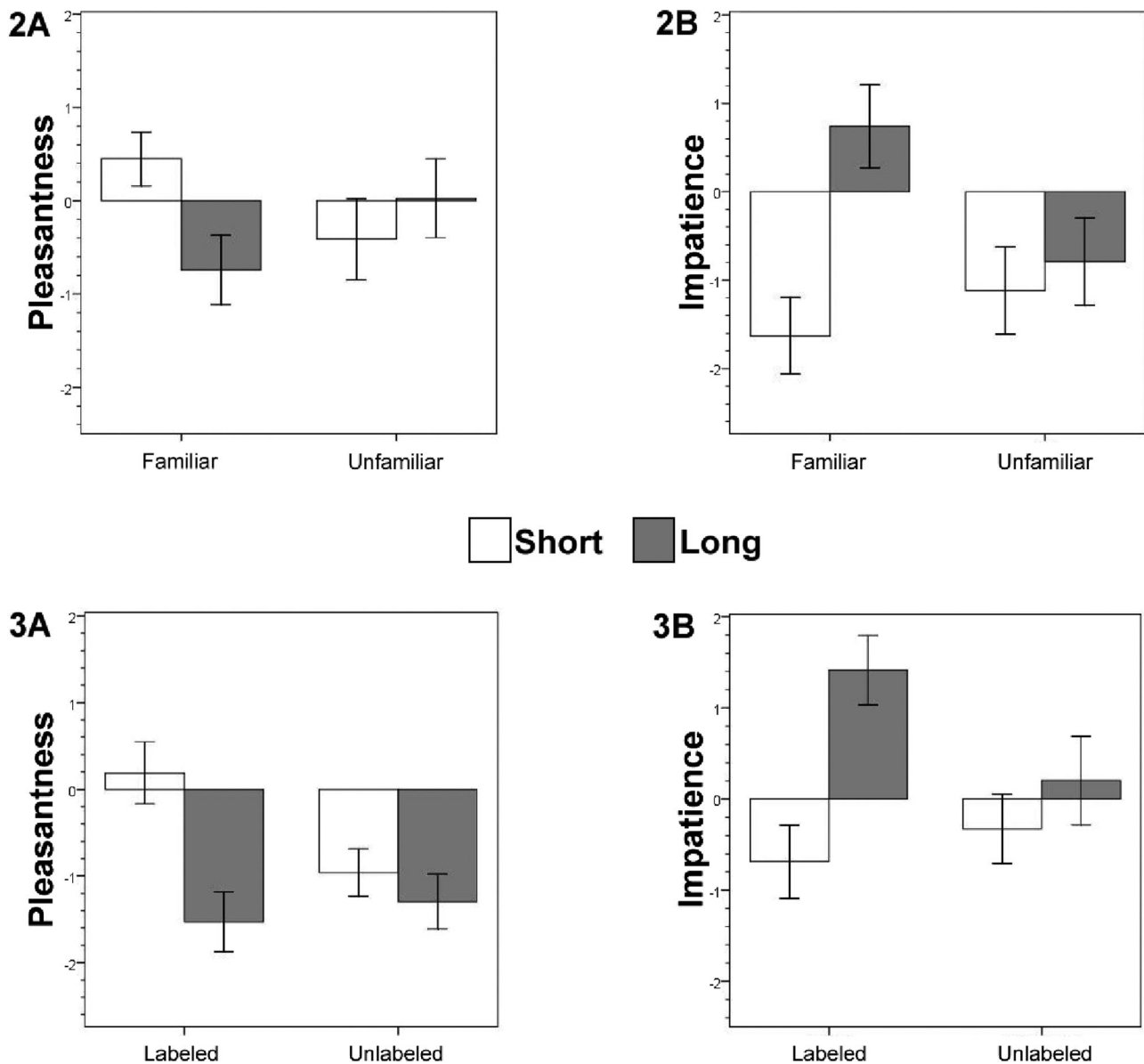


Figure 2. Participants were sensitive to the duration of a familiar sound but were not sensitive to the duration of an unfamiliar sound in Experiment 2 (2A and 2B). Participants were sensitive to the duration of a single sound when provided with a familiar label but were not sensitive to its duration when the sound was unlabeled in Experiment 3 (3A and 3B). Bars represent ± 1 standard error.

Discussion

Consistent with previous research demonstrating duration neglect (Fredrickson & Kahneman, 1993; Kahneman et al., 1993; Redelmeier & Kahneman, 1996; Schreiber & Kahneman, 2000), retrospective evaluations and affective reactions were insensitive to the duration of an unfamiliar experience (i.e., a synthesizer bleep). Indeed, participants evaluating the unfamiliar experiences may have attended most carefully to their most intense and final moments and evaluated the short and long experiences so similarly because their most intense and final moments were identical. Consistent with our hypothesis, however, retrospective evaluations and affective reactions were highly sensitive to the duration of a familiar experience (i.e., a ringing telephone). Participants evaluating familiar experiences exhibited sensitivity to a 15-s difference in the duration of sounds, despite judging those sounds in isolation. Whether prospective and retrospective evaluations made in isolation exhibit sensitivity to duration appears to depend on the familiarity of the evaluated stimulus.

Experiment 3

The previous experiments demonstrated that people are sensitive to the duration of familiar stimuli and insensitive to the duration of unfamiliar stimuli. This final experiment was designed to examine which aspect of stimulus familiarity leads evaluators to be sensitive to duration. Duration sensitivity may depend on familiarity with the stimulus itself or the cognitive accessibility of a consistent set of referents that are evoked when one perceives the stimulus to be a member of a familiar category. In other words, people may be insensitive to the duration of an unfamiliar stimulus because their attention is drawn to other more salient features of the stimulus the first time they perceive it, or people may be insensitive to the duration of an unfamiliar stimulus because it does not naturally evoke a consistent set of referents with which its duration can be compared.

In Experiment 3, all participants heard and then evaluated the same unfamiliar sound. For some, the sound was not labeled. For others, it was labeled as a stimulus belonging to a familiar category (i.e., a ringing telephone). If insensitivity to the duration of unfamiliar experiences is due to the greater salience of other features during one's first encounter with a stimulus, then participants should be insensitive to duration whether or not they were provided with a familiar label. If insensitivity to the duration of unfamiliar experiences is due to the lack of a consistent set of cognitively accessible referents, as we suggest, participants should be sensitive to duration when provided with a familiar label and insensitive to duration when not provided with such a label. The presence of the label should prompt participants to compare the sound with ringing telephones they have previously heard and thus make its duration evaluable.

Method

Participants. One hundred ninety-seven pedestrians in Harvard Square in Cambridge, Massachusetts (117 women; $M_{\text{age}} = 30.4$ years, $SD = 15.6$), agreed when asked to participate in a short experiment in exchange for candy. At a table, participants listened to an audio track via headphones and assessed that track privately.⁴

Stimuli. Participants listened to an audio track of an unusual cellular telephone ring tone on an iRiver 1-GB portable MP3 player at a consistent volume. Each ring lasted 0.8 s, with rings separated by a delay of 0.4 s.

Procedure. In a between-subjects design, participants listened to one audio track either 7.5 s in duration (i.e., short) or 20 s in duration (i.e., long). Participants provided with a label for the experience were informed that they would evaluate an Australian telephone ringing before hearing the audio track (i.e., label condition). Other participants were simply told they would evaluate an audio track before hearing it (i.e., no-label condition). After listening to the audio track, participants evaluated how pleasant the experience was and how impatient they felt during the experience on scales identical to the scales described in Experiment 2. Participants were randomly assigned to conditions.

Results

Evaluations of the pleasantness of the experiences were submitted to a 2 (duration: long, short) \times 2 (label: label, no label) ANOVA, which yielded a significant main effect of duration, $F(1, 191) = 11.13$, $p = .001$, $\eta_p^2 = .06$, which was qualified by the predicted interaction, $F(1, 191) = 5.30$, $p = .02$, $\eta_p^2 = .03$. No main effect of label was found, $F(1, 191) = 1.57$, $p = .21$, $\eta_p^2 < .01$. Planned contrasts revealed that among participants provided with a label (i.e., who were told they were listening to an Australian telephone ringing), those listening to the longer audio track considered their experience less pleasant than did those listening to the shorter audio track, $F(1, 191) = 8.98$, $p = .003$. Participants not provided with a label showed no reliable differences in their ratings of the pleasantness of the longer and shorter audio tracks, $F < 1$ (see Figure 2).

A 2 (duration: long, short) \times 2 (label: label, no label) ANOVA on the extent to which participants reported feeling impatient during the experience yielded a main effect of duration, $F(1, 191) = 10.68$, $p = .001$, $\eta_p^2 = .05$, which was qualified by the predicted interaction, $F(1, 191) = 3.69$, $p = .056$, $\eta_p^2 = .02$. Planned contrasts revealed that among participants provided with a label, participants listening to the longer audio track reported feeling more impatient during the experience than did participants who listened to the shorter audio track, $F(1, 191) = 13.86$, $p < .001$. Participants not provided with a label showed no reliable differences in the impatience they reported feeling while listening to the longer and shorter audio tracks, $F < 1$ (see Figure 2).

Discussion

Consistent with previous findings demonstrating duration neglect, retrospective evaluations of a novel stimulus were insensitive to duration when the stimulus was unlabeled and perceived as unfamiliar. Consistent with our hypothesis, however, retrospective evaluations were sensitive to the duration of a novel stimulus when it was labeled an Australian telephone ring, as participants could presumably compare the duration of the stimulus with the duration of previous familiar experiences. Indeed, participants who were

⁴ We did not include the responses of participants under the age of 15 years who completed the survey. Including those participants does not change the results of the analyses reported here.

provided with a label for the stimulus they evaluated exhibited sensitivity to a 12.5-s difference in the duration of those stimuli, despite judging them in isolation. Participants' sensitivity to the duration of an isolated stimulus thus depended on its perceived familiarity. When an unfamiliar stimulus was perceived to belong to a familiar category, its duration was evaluable.

General Discussion

Intuition suggests that the duration of stimuli should be an important determinant of their value. Yet people are often insensitive to duration, whether evaluating the pleasantness of films, sounds, and exposure to extreme temperatures or the pain experienced during invasive medical procedures (Ariely, 1998; Fredrickson & Kahneman, 1993; Kahneman et al., 1993; Redelmeier & Kahneman, 1996; Schreiber & Kahneman, 2000). The present research demonstrates that whether people are sensitive to the duration of an experience depends, in part, on the familiarity of that experience. Evaluations of a stimulus were sensitive to its duration when the stimulus was familiar or perceived as a member of a familiar category, so long as it evoked the cognitive accessibility of a consistent set of referents with which its duration could be compared (Experiment 3). When these findings are considered with the findings of previous research (Ariely & Loewenstein, 2000; Hsee et al., 1999; Liersch & McKenzie, in press; Stewart & Chater, 2003), it appears that duration neglect is not simply an attentional phenomenon. Rather, duration neglect in evaluations of unfamiliar stimuli appears to be due to a lack of present or cognitively accessible standards of comparison. People should thus neglect the duration of a stimulus that is perceived to be unfamiliar or that is actually unfamiliar, but they should be sensitive to the duration of a stimulus that is perceived to be familiar or that is actually familiar if they possess distributional information about its category or can simultaneously compare its duration with the duration of other similar stimuli.

Duration was the only scope variable examined in the experiments we report, but it is likely that the implications of their results extend to other scope variables. People are similarly insensitive to a host of other scope variables when judging unfamiliar stimuli in isolation, but they exhibit sensitivity to those same scope variables to when judging those unfamiliar stimuli with comparison to other stimuli. Weight, volume, number of lives saved, salary, cost, and quantity are but a few examples (for reviews, see Hsee et al., 1999; Hsee, Rottenstreich, & Xiao, 2005).

Assuming such an overlap exists, which preliminary data suggest (Morewedge, Kassam, Hsee, & Caruso, 2008), our results suggest three ways to instill familiarity and thus scope sensitivity in isolated judgments. One way is learning through exposure and education. Participants demonstrated acute sensitivity to duration when evaluating the pleasantness of a familiar experience. Although the present research does not identify the specific amount of familiarity with an experience that is necessary to engender duration sensitivity, repeated exposure to stimuli from a particular category or explicitly providing decision makers with not only the mean judgment but also the range and/or standard deviation should improve their assessments. Participants in the present research appeared to use implicit norms of duration in their judgments of familiar experiences in all three experiments. In this vein, Ariely and Loewenstein (2000) have demonstrated that providing deci-

sion makers with an explicit norm, a standard experience, and asking them to repeatedly choose between the standard and recent experiences also makes decision makers exhibit sensitivity to the duration of unfamiliar experiences.

Second, explicit or implicit comparison to familiar categories should lead decision makers to retrieve distributional information and consistent comparison standards from memory, increasing the reliability of evaluations and scope sensitivity (Gourville, 1998; Kahneman & Miller, 1986; Stevens, 1975). Labeling a stimulus as an example of a familiar category made its duration evaluable in isolation in Experiment 3. Presumably, participants provided with a label compared the duration of the stimulus with the duration of other phone rings previously experienced, whereas participants not provided with a label recalled comparison standards insufficiently consistent (across individual participants) to exhibit duration sensitivity.

Explicitly suggesting comparison to a familiar category appears to instill sensitivity to other scope variables. Suggesting to participants that nonmarket resources such as health can be traded just like market goods, for instance, increases the reliability of agreement between the discount rates they apply to those nonmarket resources and the discount rates they apply to money (Baron, 1997; Chapman, 2002). It is possible that if the frequent commuters in Experiment 1 were prompted to evaluate the unfamiliar route by comparing it with a route that was familiar to them, they would have demonstrated duration sensitivity. By evoking or suggesting familiar comparisons, even isolated evaluations may achieve scope sensitivity. Identifying which particular environmental and psychological factors evoke familiar comparisons, however, is a question for future research.

Third, scope sensitivity may be instilled by presenting scope values of unfamiliar stimuli in a familiar format. Presenting scope values along scales used to evaluate familiar stimuli may allow evaluators to make reliable distinctions between values of an unfamiliar scope variable by translating those values into familiar scope variables. It may be more difficult for nonexperts to determine whether a wine rated 4.5 on a 5-point scale is noticeably better than a wine rated 4.1, for example, than to determine whether a wine rated 92 on a 100-point scale is noticeably better than a wine rated 85 (Rivlin, 2006). The values along the latter scale are easier to evaluate, because one can compare wine ratings as if they were examination grades (e.g., A- and B). Provided with a familiar scale, judges may thus better discern differences between scopes of unfamiliar stimuli and more reliably rate stimuli in accordance with their value.

In addition to their implications for scope sensitivity, the results of these experiments may have important implications for psychophysical judgments of stimuli along prothetic dimensions—continuous dimensions in which an increase in the intensity of the physical stimulus results in an increase in the quantity of the psychological stimulus (e.g., loudness). Magnitude estimates typically adhere to a lawful form (Fechner, 1860/1966; Stevens, 1975; Weber, 1846/1965), but estimates more reliably adhere to such forms when raters have had previous training or experience with the evaluated stimuli (Staddon, King, & Lockhead, 1980; Thorndike & Woodworth, 1901). The greater reliability that training and experience engender may be due to the greater cognitive accessibility of a more consistent set of comparison stimuli rather than greater skill in estimating magnitudes, as some have previ-

ously suggested (e.g., Stevens, 1975). These findings may also help elucidate which kinds of comparison stimuli are likely to influence magnitude estimates. Considered together with the findings of Stewart and Chater (2003), it appears that previously experienced stimuli only influence magnitude estimations when top-down processes such as beliefs and expectations lead judges to perceive target and comparison stimuli as belonging to the same stimulus category. We would expect future research to be able to demonstrate that the loudness of a previous sound (e.g., a car alarm) should only affect judgments of a target sound (e.g., an alarm clock), for example, when judges perceive them at a categorization level broad enough to include both sounds (e.g., alarms).

Generally, the intensity of a stimulus is more evaluable and the duration of a stimulus is less evaluable. The intensity of a kiss is more evaluable than the amount of time that is spent kissing. And the intensity of momentary pain felt during a medical procedure is more evaluable than its duration because people possess an internal (psychophysical) scale to evaluate the intensity of that pain (Fernandez & Turk, 1992; Hsee, Yang, Li, & Shen, in press). Thus people often observe insensitivity to duration and rarely observe insensitivity to intensity. If people have sufficient familiarity with a stimulus, however, then its duration will also be evaluable. We would expect a veteran commuter to be more sensitive than a tourist to the relative length of a drive through her city and a spouse to be more sensitive to the length of an intimate kiss when celebrating her 25th wedding anniversary than when celebrating her first. Just as participants were sensitive to the duration of a sound when provided with a label that evoked relevant comparisons, people should be sensitive to duration when past experience, training, or expectations evoke relevant comparison information. In sum, duration neglect is not a unique phenomenon. Rather, duration neglect is a particularly important case of insensitivity to attributes that are difficult to evaluate in isolation, which can be overcome when people have the experience necessary to make appropriate comparisons.

References

- Ariely, D. (1998). Combining experiences over time: The effects of duration, intensity changes and on-line measurements on retrospective pain evaluations. *Journal of Behavioral Decision Making*, *11*, 19–45.
- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*, 157–162.
- Ariely, D., Kahneman, D., & Loewenstein, G. (2000). Joint comment on “When does duration matter in judgment and decision making?” (Ariely & Loewenstein, 2000). *Journal of Experimental Psychology: General*, *129*, 524–529.
- Ariely, D., & Loewenstein, G. (2000). When does duration matter in judgment and decision making? *Journal of Experimental Psychology: General*, *129*, 508–523.
- Ariely, D., & Zauberman, G. (2000). On the making of an experience: The effects of breaking and combining experiences on their overall evaluation. *Journal of Behavioral Decision Making*, *13*, 219–232.
- Ariely, D., & Zauberman, G. (2003). Differential partitioning of extended experiences. *Organizational Behavior and Human Decision Processes*, *91*, 128–139.
- Baron, J. (1997). Biases in the quantitative measurement of values for public decisions. *Psychological Bulletin*, *122*, 72–88.
- Chapman, G. B. (2002). Your money or your health: Time preferences and trading money for health. *Medical Decision Making*, *22*, 410–416.
- Desvousges, W. H., Johnson, F., Dunford, R., Hudson, S., Wilson, K., & Boyle, K. (1993). Measuring resource damages with contingent valuation. In J. A. Hausman (Ed.), *Contingent valuation: A critical assessment* (pp. 91–159). Amsterdam: North Holland.
- Diener, E., Wirtz, D., & Oishi, S. (2002). End effects of rated life quality: The James Dean effect. *Psychological Science*, *12*, 124–128.
- Dunn, E. W., Wilson, T. D., & Gilbert, D. T. (2003). Location, location, location: The misprediction of satisfaction in housing lotteries. *Personality and Social Psychology Bulletin*, *29*, 1421–1432.
- Fechner, G. T. (1966). *Elements of psychophysics* (H. Adler, Trans.). New York: Holt, Rinehart & Winston. (Original work published 1860)
- Fernandez, E., & Turk, D. C. (1992). Sensory and affective components of pain: Separation and synthesis. *Psychological Bulletin*, *112*, 205–207.
- Fredrickson, B. L. (2000). Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition & Emotion*, *14*, 577–606.
- Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology*, *65*, 45–55.
- Gourville, J. T. (1998). Pennies-a-day: The effect of temporal reframing on transaction evaluation. *Journal of Consumer Research*, *24*, 395–408.
- Hsee, C. K. (1998). Less is better: When low-value options are valued more highly than high-value options. *Journal of Behavioral Decision Making*, *11*, 107–121.
- Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin*, *125*, 576–590.
- Hsee, C. K., Rottenstreich, Y., & Xiao, Z. (2005). When more is better: On the relationship between magnitude and subjective value. *Current Trends in Psychological Science*, *14*, 234–237.
- Hsee, C. K., Yang, Y., Li, N., & Shen, L. (in press). Wealth, warmth, and wellbeing: Whether happiness is relative or absolute depends on whether it is about money, acquisition or consumption. *Journal of Marketing Research*.
- Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science*, *4*, 401–405.
- Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, *93*, 136–153.
- Kahneman, D., Ritov, I. & Schkade, D. (1999). Economic preferences or attitude expressions? An analysis of dollar responses to public issues. *Journal of Risk and Uncertainty*, *19*, 220–242.
- Kahneman, D., Wakker, P. P., & Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *Quarterly Journal of Economics*, *112*, 375–405.
- Laming, D. R. J. (1997). *The measurement of sensation*. London: Oxford University Press.
- Liersch, M. J., & McKenzie, C. R. M. (in press). Duration neglect by numbers—and its elimination by graphs. *Organizational Behavior and Human Decision Processes*.
- Morewedge, C. K., Gilbert, D. T., Keysar, B., Berkovits, M. J., & Wilson, T. D. (2007). Mispredicting the hedonic benefits of segregated gains. *Journal of Experimental Psychology: General*, *136*, 700–709.
- Morewedge, C. K., Gilbert, D. T., & Wilson, T. D. (2005). The least likely of times: How remembering the past biases forecasts of the future. *Psychological Science*, *16*, 626–630.
- Morewedge, C. K., Kassam, K. S., Hsee, C. K., & Caruso, E. M. (2008). [Estimates and relations of judgments of familiarity and length]. Unpublished raw data, Carnegie Mellon University.
- Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, *72*, 407–418.
- Parducci, A. (1995). *Happiness, pleasure, and judgment: The contextual theory and its applications*. Mahwah, NJ: Erlbaum.

- Redelmeier, D. A., & Kahneman, D. (1996). Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain, 66*, 3–8.
- Rivlin, G. (2006, August 13). Wine ratings might not pass the sobriety test. *The New York Times*, p. 1.
- Roy, M. M., Christenfeld, N. J. S., & McKenzie, C. R. M. (2005). Underestimating the duration of future events: Memory incorrectly used or memory bias? *Psychological Bulletin, 131*, 738–765.
- Schreiber, C. A., & Kahneman, D. (2000). Determinants of the remembered utility of aversive sounds. *Journal of Experimental Psychology: General, 129*, 27–42.
- Staddon, J. E. R., King, M. C., & Lockhead, G. R. (1980). On sequential effects in absolute judgment experiments. *Journal of Experimental Psychology: Human Perception and Performance, 6*, 290–301.
- Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New York: Wiley.
- Stewart, N., & Chater, N. (2003). No unified scales for perceptual magnitudes: Evidence from loudness. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th annual conference of the Cognitive Science Society* (pp. 1116–1121). Boston, MA: Cognitive Science Society.
- Stutzer, A., & Frey, B. S. (2004). *Stress that doesn't pay: The commuting paradox* (IZA Discussion Paper No. 1278). Bonn, Germany: Institute for the Study of Labor (IZA).
- Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions: II. The estimation of magnitudes. *Psychological Review, 8*, 384–395.
- Varey, C., & Kahneman, D. (1992). Experiences extended across time: Evaluation of moments and episodes. *Journal of Behavioral Decision Making, 5*, 169–185.
- Weber, E. H. (1965). On the sense of touch and common sensibility. In B. Haupt (Trans.) & R. J. Herrnstein & E. G. Boring (Eds.), *A sourcebook in the history of psychology* (pp. 34–39). Cambridge, MA: Harvard University Press. (Original work published 1846)
- Wirtz, D., Kruger, J., Scollon, C. N., & Diener, E. (2003). What to do on spring break? The role of predicted, on-line, and remembered experience in future choice. *Psychological Science, 14*, 520–524.
- Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience, 5*, 495–501.
- Zakay, D., & Block, R. A. (1997). Temporal cognition. *Current Directions in Psychological Science, 6*, 12–16.

Received December 12, 2007

Revision received November 12, 2008

Accepted November 19, 2008 ■

Low Publication Prices for APA Members and Affiliates

Keeping you up-to-date. All APA Fellows, Members, Associates, and Student Affiliates receive—as part of their annual dues—subscriptions to the *American Psychologist* and *APA Monitor*. High School Teacher and International Affiliates receive subscriptions to the *APA Monitor*, and they may subscribe to the *American Psychologist* at a significantly reduced rate. In addition, all Members and Student Affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

Essential resources. APA members and affiliates receive special rates for purchases of APA books, including the *Publication Manual of the American Psychological Association*, and on dozens of new topical books each year.

Other benefits of membership. Membership in APA also provides eligibility for competitive insurance plans, continuing education programs, reduced APA convention fees, and specialty divisions.

More information. Write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242.