Journal of Behavioral Decision Making, Vol. 11, 19–45 (1998)

Combining Experiences Over Time: The Effects of Duration, Intensity Changes and On-Line Measurements on Retrospective Pain Evaluations

DAN ARIELY*

The Fuqua School of Business, Duke University, USA

ABSTRACT

Two experiments were conducted to examine the effects of various factors on retrospective pain evaluation. The factors examined in Experiment 1 were the rate and pattern of change, the intensity (particularly the final intensity), and the duration of the painful experience. Experiment 2 manipulated these factors and, in addition, examined the effect of continuous (on-line) ratings on the overall retrospective evaluation. The two experiments utilized different pain modalities, heat in the first and mechanical pressure in the second. In addition, all subjects in Experiment 1 experienced stimuli with the same physical magnitude, while in Experiment 2 stimuli were individually tailored to make them subjectively equivalent. In both experiments, subjects were presented with a series of painful stimuli and evaluated the intensity of each stimulus immediately upon its termination. The stimuli themselves were composed of multiple intensity levels that differentially changed over time (Intensity-Patterns). Subjects' on-line ratings in Experiment 2 closely mirrored the physical patterns of the intensities. The main conclusion from both experiments is that the retrospective evaluations of painful experiences are influenced primarily by a combination of the final pain intensity and the intensity trend during the latter half of the experience. In addition, results indicated that duration has little impact on retrospective evaluations for stimuli of relatively constant intensity. However, when the stimulus intensity changes over time, duration does play a role. Finally, the task of continuously reporting the stimulus intensity had a moderating impact on the retrospective evaluations. © 1998 John Wiley & Sons, Ltd.

Journal of Behavioral Decision Making, **11:** 19–45, 1998.

KEY WORDS: pain evaluation; combining experiences; time; retrospective evaluations

* Correspondence to: Dan Ariely, The Fuqua School of Business, Duke University, Box 90120, Durham, NC 27708-0120, USA. e-mail: ariely@mail.duke.edu.

CCC 0894–3257/98/010019–27$17.50
© 1998 John Wiley & Sons, Ltd.

Received 19 December 1995
Accepted 5 March 1997

INTRODUCTION

Most experiences we have in our day-to-day lives extend over a certain time duration, be it a few seconds or many hours. Within this duration, the intensity of experiences is seldom constant and its magnitude usually changes over time. A medical treatment, for example, is most likely associated with some moments that are very unpleasant and some that are less so (i.e. a change in intensity over time). Once such an experience is over, one can form an overall evaluation of it, capturing the remembered intensity of the experience as a whole. Since this global evaluation is likely to play an important role in future behavior and decisions, it is important to understand the ways in which this evaluation is constructed. The current work is aimed at understanding the relationship between the momentary intensity over time and the global evaluation for the experience as a whole. This is done by examining the impact different aspects of the ongoing experience have on the construction of the overall evaluation. Although this work examines only one domain of experience (namely pain), one can speculate that the relationship between momentary and overall evaluations will apply to other domains as well.

How do changes in pain intensity and duration interact to affect pain perception and its subsequent memory? Despite good practical reasons for answering this question (Hunter, Philips and Rachman, 1979; Jamison, Sbrocco and Winston, 1989), little is known about the topic. For example, almost all the nurses and doctors I have met believe that the best way to remove dressings from burn patients is to do so quickly, with fast, strong motions (see also Choiniere *et al*., 1990; Perry, 1984). In order to check my personal impressions, I conducted a small survey regarding this question. Eighteen members of the medical staff in a large burn center participated in this survey and were asked for their treatment method and preference. The results indicated that a large majority (89%) believed that increasing pain intensity and shortening the duration decreases overall remembered pain. The belief in this method is strong, although it is clear that it causes more intense momentary pain. The general wisdom is that by minimizing pain duration, overall pain, as well as its remembered intensity, are also minimized. As a long-time burn patient, I suspect that these principles are wrong, but little guiding research is available.

Most people occasionally experience bodily pain in their daily lives. These painful experiences extend over various lengths of time, during which the intensity usually fluctuates from moment to moment rather than being constant. Changes in pain intensity occur not only for pain inflicted upon us by other people, such as dentists, physiotherapists or orthopedists, but also for internally generated pain, such as headaches and backaches. These and other changes in pain intensity over time can be seen as a 'pattern of pain' in which intensity is plotted as a function of time. Henceforth these patterns of pain will be referred to as the *Intensity-Patterns*. Therefore, the goal of the current work is to map the joint effects of changes in pain intensity (Intensity-Patterns) and duration on subsequent memory for pain.

Although it is clear that pain intensity fluctuates over time, it is unclear whether the pattern of fluctuation affects the perceived intensity of the entire painful experience. It is worth noting that much of the psychological research (at least implicitly) assumes that the perception of an experience is at least monotonically related to the perception of its components. Recently, there has been a growing interest in the monotonicity assumption as well as in rules governing the combination of experiences (Fredrickson and Kahneman, 1993; Kahneman *et al*., 1993; Redelmeier and Kahneman, 1993; Varey and Kahneman, 1992; Loewenstein and Prelec, 1993). In an insightful paper, Loewenstein and Prelec (1993) looked into this question by using sequences of experiences, and demonstrated that the ordering of experiences in time strongly influences the overall evaluation. This result is important because it demonstrates that the rules for combining experiences are sensitive to their sequence in time (i.e. to the pattern). Referring back to our question, it suggests that the patterns of fluctuations affect overall pain evaluations.

In addition, several researchers have started to examine the relationship between global retrospective evaluations of experiences and on-line (i.e. moment-by-moment) evaluations of the same experiences.

 © 1998 John Wiley & Sons, Ltd.

In this way, it is possible to examine joint effects of duration and subjective patterns of the experience on the overall retrospective evaluation. In an interesting study, Varey and Kahneman (1992) tested an hypothesis stating that experience over time is not a simple combination of its discrete components, but is influenced primarily by some particular aspects of the stimulus. Results from their experiments as well as others (Fredrickson and Kahneman, 1993; Kahneman *et al.*, 1993; Redelmeier and Kahneman, 1993) suggest that the passage of time (i.e. prolonging the experience) does not affect the subsequent overall judgment. Insensitivity to duration is a counter intuitive result, since it is commonly thought that prolonging the duration of a painful experience increases its overall perceived pain. In a simple additive way, one would expect the overall evaluation of a painful experience to cumulate with time, which is exactly what Varey and Kahneman (1992) showed not to occur.

Fredrickson and Kahneman (1993) named this phenomenon 'duration neglect'. They offered an appealing metaphor borrowed from Milan Kundera (Kundera, 1991), according to which 'memory does not take film, it takes photographs'. This metaphor is used to express the idea that memory is built not on continuous experience but only on some selected key aspects of the experience remembered. In line with this metaphor, the results from their study, as well as the results by Kahneman *et al.* (1993), suggest that the overall evaluation of an experience is best predicted by a weighted average of the most extreme part of the experience (its peak) and its final (end) intensity, regardless of the experience duration. It is important to note that this conclusion is not without caveats. For example, Carmon and Kahneman (1996) found that in a simulated queuing experience, when subjects judged moment-by-moment as well as overall satisfaction, the final, but not the peak intensity level, had an effect on the overall satisfaction. Returning to the domain of pain, Price and Tursky (1975) have shown that subjects are sensitive to the passage of time when presented with changing levels of painful stimuli.

A different line of research, dealing with many of the same general issues, involves preferences for experiences that change over time. Looking at the questions of changes in outcomes and changes in levels of satisfaction, Hsee and Abelson (1991; see also Hsee, Abelson and Salovey, 1991) suggest that the rate of outcome change (i.e. its slope) is the determining factor for the overall global evaluations. An extension of this work by Loewenstein and Prelec (1993) demonstrated that when dealing with sequences of events, subjects have a preference for improving trends (i.e. declining sequences of pain). In one of their examples, subjects were asked to choose whom they would visit first, an annoying aunt or a good friend. The results showed that there was a strong preference to visit the annoying aunt first, savoring the visit of the friend for later. This again implies that the judgment of an overall experience is not simply composed of its independent discrete aspects, but rather it is influenced by the relationships among them. More specifically, these judgments are primarily influenced by the final intensity trend of the experience.

The importance of changes over time has been established not only in behavioral decision research but also in research on visual and other perceptual systems. For example, it is suggested that perception of time duration is influenced by the amount of change within the interval being judged (Avant, Lyman and Antes, 1975; Poynter and Holma, 1985). In a different domain, it has been shown that eliminating change has strong implications for the visual system. One such example uses a presentation of a stable retinal image to the eye. Subjects who are exposed to these stimuli report that the visual characteristics of the image fade rapidly, making it disappear (Pritchard, Heron and Hebb, 1960; Yarbus, 1967). For these reasons, it is argued that changes are crucial to the perceptual system. Moreover, changes are said to be crucial because of the way perceptual sensitivity changes in reaction to the intensity levels of the external stimuli. This change in sensitivity has been called adaptation. It can be claimed that adaptation has both positive and negative sides to it (for more details, see Helson, 1964). On the negative side, adaptation does not allow one to estimate the absolute intensity of the stimuli or to remain at the same sensitivity level when exposed to a prolonged stimulus. On the positive side, adaptation assures great sensitivity to changes within the range of a given adaptation level.

© 1998 John Wiley & Sons, Ltd.

Kundera's metaphor (Kundera, 1991) suggests that the integration of information over time cannot be perfect nor complete, and therefore some of the information is not utilized. In addition, it is likely that the perceptual system is differentially sensitive to some particular aspects of the stimulus structure, such as its intensity change. From an adaptation point of view, then, it seems that there will be a profound difference between stimuli that have constant levels of intensity and stimuli with changing levels of intensity.

The current research uses stimuli of moderate to strong intensity to explore the relationship between physical changes in the levels of the painful stimuli and the overall retrospective evaluation of the whole painful episode. The main question addressed is what factors influence and determine global pain evaluation. Candidate factors include physical intensity at the end, beginning, and the peak of the episode; the duration of the episode; the trend (Loewenstein and Prelec, 1993); or perhaps, as suggested by Hsee and Abelson (1991), the rate of intensity change. The two experiments presented next utilize two different modalities for pain elicitation: heat and mechanical pressure, respectively. Aside from the difference in pain modality, these two experiments differ in two design features. Subjects in the first experiment all experienced stimuli of the same physical intensity, while for those in the second, stimuli were individually tailored to fit each individual subject's pain threshold. A second important difference in the two experiments is that in Experiment 1, measurements consisted only of retrospective evaluations which were analyzed as a function of the physical intensity of the stimuli. In Experiment 2, subjects also reported continuous intensity evaluations throughout the painful experience. Therefore, data from Experiment 2 were analyzed as a function of both the physical intensity and of the reported subjective intensity.

EXPERIMENT 1

The experimental hypotheses of Experiment 1 are:

H1: As suggested by Kahneman *et al.* (1993), the final intensity of an experience impacts the overall retrospective evaluation.

H2: In line with Loewenstein and Prelec (1993), the final trend of the experience plays a large role in the overall retrospective evaluation.

H3: Stemming from the ideas of Hsee *et al.* (1991), rate of intensity change influences pain perception and memory such that when the rate of change is high and positive, pain ratings will be higher.

H4: Finally, in line with Helson's (1964) adaptation level theory, duration is hypothesized to have a smaller role in the perception of stimuli with constant intensity levels compared with the perception of stimuli that have changing intensity levels.

Method

Subjects

Twenty subjects, mainly graduate students and faculty from the University of North Carolina and Duke University, participated voluntarily in a one-hour pain perception experiment. Subjects were 19 to 55 years old and all in good health. Five of the subjects were female.

Stimuli

The stimuli consisted of precisely controlled applications of heat by means of a 100 mm^2 contact Thermode placed on the inner part of the forearm. The heating element was computer controlled and

 © 1998 John Wiley & Sons, Ltd.

was actively cooled through the use of refrigerated anti-freeze fluid. The warming and cooling slopes of this Thermode are limited to approximately 10°C/s. All the stimuli were administered as increments over a baseline of 35°C (95°F), so that the heating element was at a temperature of at least 35°C at all times. Locations on the left and right arms were used interchangeably at the discretion of the subject.

Intensity-Patterns over time were manipulated within two distinct within-subject factorial designs. The first design, called the *constant design*, consisted of two factors: Duration and Intensity (temperature). Duration had two levels (10 and 14 seconds), and Intensity had four levels (45°C, 46°C, 47°C, 48°C which are 113°F, 114.8°F, 116.6°F and 118.4°F, respectively), thus composing a total of eight distinct stimuli in the constant design (see Exhibit 1B). The second design, called the *patterned design*, consisted of two factors as well: Time-Relationship and Intensity-Patterns. Time-Relationship had three levels and Intensity-Patterns had eight levels. The patterned design is schematically represented in Exhibit 1A. (Note the eight Intensity-Pattern traces in column A and their labels.) The stimulus labeled Low&Up is a stimulus that was 45°C for the first half of the duration and then sloped up to 48°C. The next stimulus, called Down&Low, is a vertical reflection of the first one, starting from 48°C, sloping down to 45°C at the midpoint of its duration, and then remaining constant at 45°C until the end. All the remaining Intensity-Patterns can be interpreted similarly. Note that the first four stimuli and the next two pairs are each mirror images of one another.

Due to constraints imposed by the heating device, Intensity-Patterns were not accelerated smoothly, instead they were composed of constant temperature segments, as shown in columns B through D. The three Time-Relationships were determined by the length of the segments. In column B the seven heat segments were 1.43 seconds long and the full stimulus always lasted 10 seconds. In column C the 7 heat segments were 2 seconds long and the full stimulus always lasted 14 seconds. Because the Intensity-Patterns in column C were a 'stretched' version of the Intensity-Patterns in column B, the Intensity-Patterns differed not only in duration but also in rate of temperature change. Therefore, column D was created to equal column C in duration and column B in rate of change. This was accomplished by replicating column B Intensity-Patterns but adding 2 seconds at the beginning and at the end of each interval, so that the full stimulus always lasted 14 seconds.

Overall, the patterned design consisted of eight different stimulus traces, each in three different Time-Relationship (columns B-D), making a total of twenty-four distinct stimuli.

Procedure

Subjects sat in an armchair next to the Thermode, and were presented with a drawing of a thermometer-like visual analog scale ranging from 0 to 100, with numerical intervals every 10 points. The scale was anchored at 0 (no pain at all) and at 100 (pain as bad as it can be). This kind of visual analog scale has been shown to be reliable across different groups of subjects including both pain patients and pain-free volunteers (Price, McHaffie and Stein, 1992). The probe of the Thermode was given to the subjects, who were asked to place the end of the probe on their forearm and to change its location after each trial. On each trial, after subjects indicated they were ready, the experimenter simultaneously started the stimulus and announced that he was doing so. The end of each trial was also noted by the experimenter, at which point subjects assessed the overall pain they had experienced during the entire trial. This assessment was based on the visual analog scale and given as a numeric-oral response to the experimenter. Seven trials were allowed as practice before the actual experiment started.

Following the initial practice, subjects experienced a total of 64 stimuli during the actual experiment. The experiment consisted of two blocks which were identical for any given subject. Each block contained the full range of stimuli, from both of the designs (8 from the constant design and 24 from the patterned design, making a total of 32 stimuli). For the presentation order of the stimuli within a block, three different random orders were created, and each subject was assigned to one of the three random

© 1998 John Wiley & Sons, Ltd.

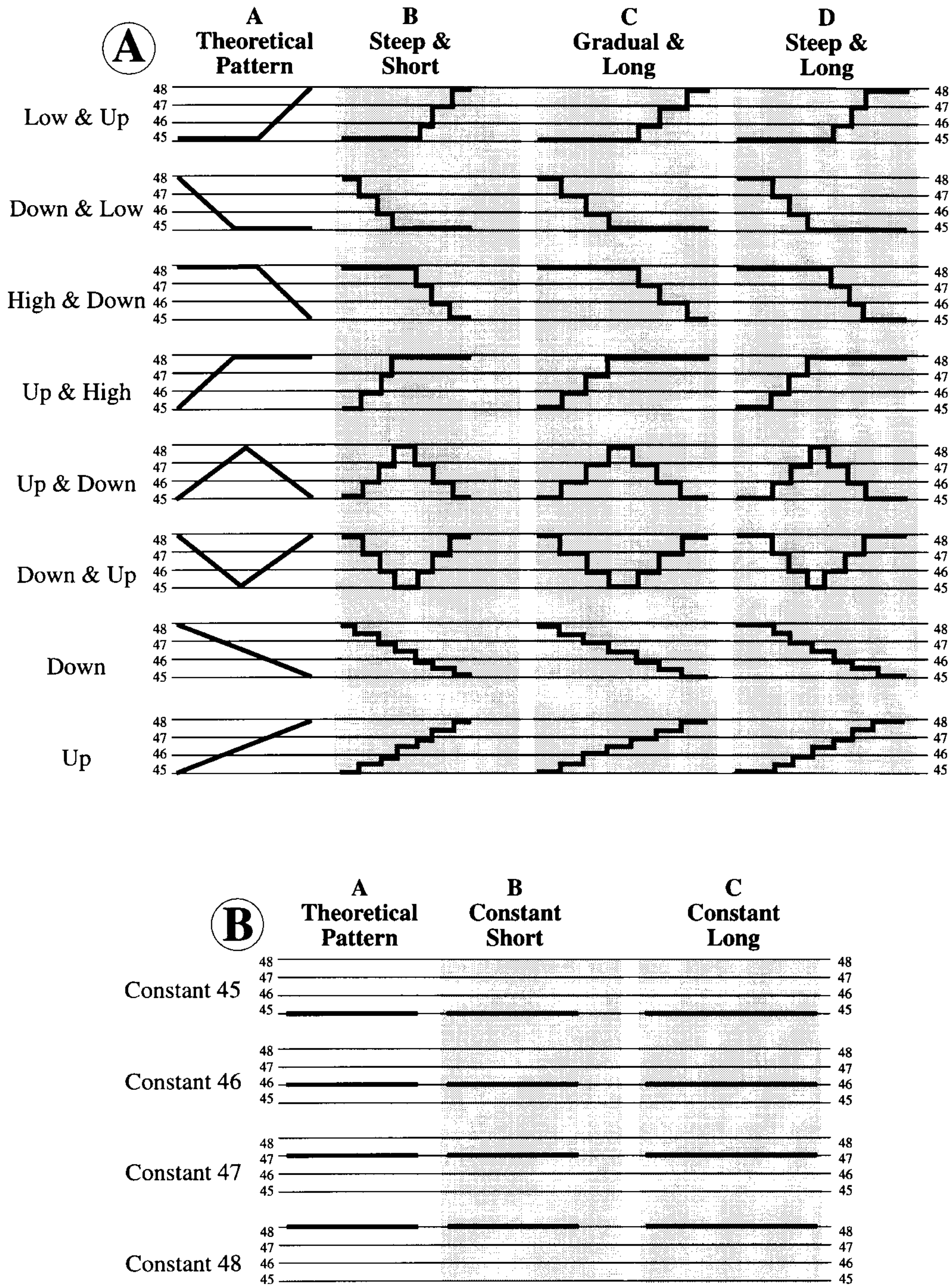

Exhibit 1. Relationship of changes in temperature over time for the different Intensity-Patterns in Experiment 1. The top panel (A) represents the stimuli of the pattern design with the theoretical relationship shown in column A and the three Time-Relationships shown in columns B, C and D. The bottom panel (B) represents the stimuli of the constant design with the theoretical relationship shown in column A and the two Time-Relationships shown in columns B and C. Time is represented on the horizontal axis and temperature in degrees Celsius on the vertical axis

© 1998 John Wiley & Sons, Ltd.

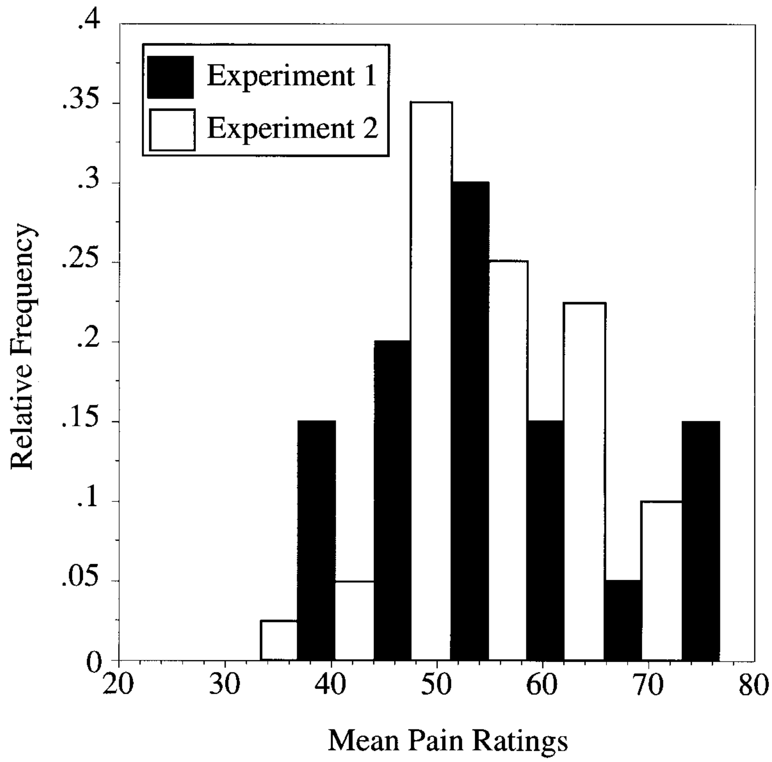

Exhibit 2. Mean response of each of the subjects on the visual pain analog over all 64 stimuli in Experiment 1 and over all 42 stimuli in Experiment 2

orders. This method was chosen in order to avoid the effect of any specific order, while making it possible to look into the question of order effects.

Results

Overall pain ratings, collapsed across subjects, had a mean of 53.0 and a standard deviation (SD) of 22.4 with a large between-subject variability in the use of the scale (see Exhibit 2). For the purpose of statistical analysis and in order to overcome individual differences in the use of the pain scale, responses for each subject were normalized. All responses of each subject were taken as a unit and each response in this unit was converted to T scores[1] (mean 50 and standard deviation 10). All subsequent analyses were performed on these converted scores.

Results for Experiment 1 are presented in the following order. First the question of presentation order for both the between and the within aspects will be discussed, followed by a discussion of the more substantive results. These results will be discussed separately for the constant and the patterned designs.

[1] Letting X be a response, $\bar{X}$ the mean of the subject's 64 responses, and S their standard deviation, each response, X_i, was converted to T by $T = 50 + 10(X_i - \bar{X})/S$.

© 1998 John Wiley & Sons, Ltd.

As mentioned earlier, each subject experienced two identical blocks with 32 painful stimuli in each. In addition, three different random orders were used for the order of the stimuli within each block. Looking first at the effect of presentation order, a four-way ANOVA was carried out, using order as a between-subject factor and block, duration, and Intensity-Pattern as within-subject factors. Presentation order had no significant main effect, nor did it interact with any of the other factors and was therefore ignored in all subsequent analyses.

Next, a three-way within-subject ANOVA was carried out separately for each of the two designs, Block by Time-Relationship by Intensity-Pattern for the patterned design, and Block by Duration by Intensity for the constant design. The constant stimulus design yielded a significant interaction between Intensity and Duration (MSE = 180, $F_{(3,57)} = 3.45$, $p = 0.022$), and a significant main effect of Intensity (MSE = 8174, $F_{(3,57)} = 111.61$, $p < 0.001$). There was no main effect for Duration, or Block, nor were there any interactions with Block. The ANOVA on the patterned design revealed a significant interaction for Time-Relationship and Intensity-Patterns (MSE = 170, $F_{(14,266)} = 2.372$, $p = 0.004$) and main effects of Intensity-Patterns (MSE = 5710, $F_{(7,133)} = 66.919$, $p < 0.001$), and Time-Relationship (MSE = 930, $F_{(2,38)} = 10.958$, $p < 0.001$). Again, no significant main effect or interaction appeared for the repeated Blocks. Looking at the order and Block factors for both designs, it is clear that subjects' responses were not affected by the specific order of the stimuli nor were they affected by the repeated exposure to pain during the experiment. Therefore, these order factors are eliminated from all future analyses.

The goal of the current experiment was to examine the role of duration, rate of change and the final intensity on encoding and remembering pain. I now explain and expand on the experimental findings in terms of these factors.

Effects of the final intensity and trend

The patterned design contained four pairs of stimuli in which each member in each pair was a rotation by 180° around the middle vertical axis of the other item in the pair (see Exhibit 1A). For the set of stimuli chosen for this experiment, this rotation kept the total heat energy for the two stimuli in a pair equal. The main difference between the two stimuli in a pair was that one had a high ending and the other a low ending. As can be seen from the left three columns of Exhibit 3, Intensity-Patterns with the

Exhibit 3. Mean converted estimated pain scores for each of the stimuli in the experiment, ranked by overall subjective intensity

Pattern's name	Steep & Short	Gradual & Long	Steep & Long	Constant Short	Constant Long	Mean Pain
Up&High	58.0	60.1	57.3	—	—	58.5
C48	—	—	—	58.4	58.3	58.3
Up	56.3	55.7	59.8	—	—	57.2
Down&Up	52.8	53.2	58.2	—	—	54.7
Low&Up	51.5	52.8	57.5	—	—	54.2
High&Down	52.0	54.6	51.9	—	—	52.8
C47	—	—	—	50.7	52.4	51.5
Up&Down	47.6	49.4	49.6	—	—	48.9
Down	43.1	42.8	45.4	—	—	43.8
C46	—	—	—	42.2	41.3	41.7
Down&Low	34.9	36.7	44.0	—	—	38.5
C45	—	—	—	32.7	38.5	35.6
Overall mean	49.5	50.7	53.0	46	47.6	49.6

 © 1998 John Wiley & Sons, Ltd.

high final intensity were rated on average as more aversive than ones with low final intensity in all three Time-Relationship conditions (overall, $\Delta\bar{X} = 10.13$, $F_{(1,266)} = 288.56$, $p < 0.001$). Identical results emerged when each of the four pairs was analyzed separately, all supporting Hypothesis 1.

Final intensity and trend effects (Kahneman *et al.*, 1993; Loewenstein and Prelec 1993) can also be seen clearly in Exhibit 3. Equating for overall intensity, not only is the member of each mirror-image pair of stimuli with the higher final intensity the more painful one, but the pair-members are in separate halves of the rankings, thus supporting Hypothesis 2. Note, however, that in this experiment a high final intensity also means a positive final slope, and conversely, a low final intensity means a negative final slope. We do not yet have the data to tease apart the separate effects of trend, on the one hand, and final stimulus intensity, on the other. But trend as well as slope effects cannot be ignored, as evidenced by the increase in pain estimation between Gradual&Long versus Steep&Long (more on this point later). Moreover, the verbal protocols of some subjects, who reported fear of strong pain with increasing intensity and relief with decreasing heat, attest to the importance of rate and direction of intensity change (also see Arntz, Van Eck and de Jong, 1991; Stevenson, Kanfer and Higgins, 1984).

Duration in constant design stimuli

Because of the Intensity by Duration interaction, four contrast comparisons were performed, one for each of the constant temperatures (i.e. 45°C, 46°C, 47°C, and 48°C). The duration effect was significant only for the 45°C stimulus ($\Delta\bar{X} = 5.8$, $F_{(1,57)} = 12.9$, $p < 0.001$), but not at any of the other temperatures (the results are shown in Exhibit 4).

It can be seen that perceived overall intensity increased with duration for the 45°C stimulus, but not for the three hotter temperatures. Why does duration matter only at the lowest temperature? One could speculate that in the constant design, the initial perception of the 45°C stimulus was that of heat only and not of pain (see Hardy, Wolff and Goodell, 1952; Price, McHaffie and Stein, 1992). Heat-induced

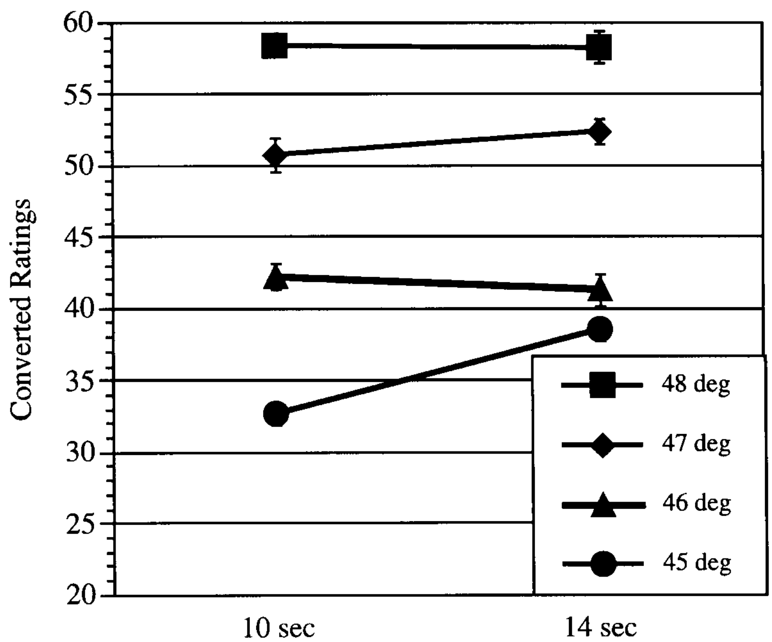

Exhibit 4. Mean converted scores for the constant design for the four temperatures and the two durations. Error bars are based on standard error

pain is experienced only once a certain temperature, duration, or a combination of both is exceeded. At that point, a threshold is reached and the sensation is labeled as pain and not just as heat. It seems that the results of the 45°C stimulus could be explained with this argument.

From looking at the individual data, it is apparent that some of the subjects did not reach a pain threshold with 10-second stimuli at 45°C (five ranked this stimulus no higher than 10), whereas all subjects reached a pain threshold with 14-second stimuli at 45°C. Therefore, the overall conclusion from the constant design is that for constant stimuli duration does not play a rule in the global evalution, thus supporting the first part of Hypothesis 4.

Duration in patterned stimuli

Turning to the patterned design, if duration has an effect under changing levels of pain intensity over time (Hypothesis 4), then we would expect the two long time relationships (columns C and D in Exhibit 1A) to be perceived as more aversive than the short Time-Relationship (column B). In addition, if there is merit to the hypothesis that the rate of change is an important factor (Hypothesis 3), then we would expect the Steep&Long Time-Relationship (column D) to be perceived as more aversive than the Gradual&Long Time-Relationship (column C).

In a pairwise contrast-comparison of the Time-Relationships, all pair comparisons were significantly different, thus supporting both Hypotheses 3 and 4: Steep&Short versus Gradual&Long ($\Delta\bar{X} = 1.274$, $F_{(1,38)} = 3.06$, $p = 0.022$); Gradual&Long versus Steep&Long ($\Delta\bar{X} = 2.103$, $F_{(1,38)} = 8.33$, $p = 0.002$); Steep&Short versus Steep&Long ($\Delta\bar{X} = 3.377$, $F_{(1,38)} = 21.48$, $p < 0.001$). As can be seen in Exhibit 5, results indicate that for the patterned design, an increase in duration as well as an increase in the rate of temperature change increase the perceived intensity of the painful episode. It is important to note that the manipulation of the rate of change in the Time-Relationship was relatively

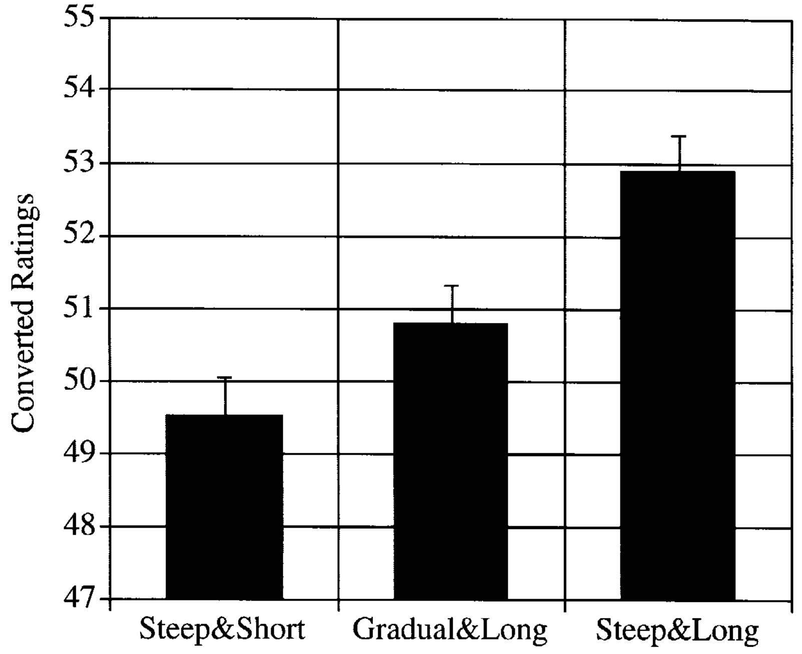

Exhibit 5. Mean converted scores for the patterned design for the different Time-Relationships. Error bars are based on standard error

© 1998 John Wiley & Sons, Ltd.

subtle and small in magnitude. Therefore, the fact that this relatively subtle manipulation had an effect implies that the importance of the slope can be much larger in other circumstances. The conclusion therefore is that the rate of change has an impact on retrospective evaluations, supporting Hypothesis 3. In addition, although duration was found not to play a role for the constant stimuli it did play a role in the pattern design, thus supporting Hypothesis 4.

However, it is also clear from the significant Time-Relationship by Intensity-Pattern interaction noted previously that this effect was not consistent over all Intensity-Patterns. Exhibit 3 shows the effect to be reversed to a small degree for the Up&High Intensity-Pattern with respect to duration and for the High&Down Intensity-Pattern with respect to slope. In all other cases, the ratings are either unaffected or increase with both duration and slope.

In addition, it is instructive to consider stimuli from both the constant and the patterned designs. For this purpose, the stimuli in Exhibit 3 have been arrayed from bottom to top according to their mean ratings over all Time-Relationship conditions (as shown at the far right side of Exhibit 3). Note first that the Up&High and constant 48° stimuli were perceived to be almost equally painful. This result could be explained if one considers the effect of the final experience to be the most important influence on pain ratings. A different direction of explanation involves adaptation to the constant stimuli, and an increased sensitivity due to the positive slope of the patterned stimuli (Loewenstein and Prelec, 1993).

Discussion

Four hypotheses were stated regarding the impact different aspects of the painful stimuli had on retrospective evaluations. These factors were: the final intensity level, the trend of the experience, the velocity of the experience and the interaction of those factors with the duration of the experience. The two designs within the experiment were constructed to shed some light on the questions that motivated the different hypotheses.

(1) Does the final experience impact the overall retrospective evaluation? In other words, is there a recency effect (Miller and Campbell, 1959)?
(2) Are these evaluations influenced by the trend of the experience?
(3) Are these evaluations influenced by the rate of change?
(4) Finally, is the sensitivity to duration decreased under constant compared with patterned stimuli?

The analysis of the different stimuli in the patterned design demonstrates that end intensity as well as the final trend play an important role in global pain evaluations. As was shown in Exhibit 3, the intensity patterns with the higher endings were rated higher than the ones with the lower endings. In addition, the rate of change had an effect on the overall retrospective evaluations. That is, equating for duration, ratings were significantly higher for the Steep&Long than for the Gradual&Long Intensity-Patterns (again, except for a single case).

The constant design showed no effect of duration at temperatures of 46°C, 47°C, and 48°C, but a strong effect at 45°C. As mentioned earlier, this could be attributed to a pain threshold (see also Cabanac, 1971). Thus, results from the constant design support the conclusions of Kahneman *et al.* (1993) that duration does not affect retrospective pain evaluations, but with the caveat that the intensity–time combination must exceed a pain threshold.

Finally, there was a small but significant difference between the Steep&Short and Gradual&Long Intensity-Patterns, suggesting a trade-off between rate of change and duration. That is, with the values used in this experiment, the increased evaluation in going from a 10- to a 14-second stimulus more than offset the decrease in going from a steep to a gradual rate of change. Overall, therefore, the conclusion is clear: with patterned stimuli, retrospective evaluations increase with both duration and rate of change.

© 1998 John Wiley & Sons, Ltd.

Why should duration matter in the patterned, but not the constant case? One possible explanation concerns adaptation (Dar, Ariely and Frenk, 1995; Helson, 1964). From this point of view, it is likely that adaptation occurred only for the constant temperature stimuli, which results in a subjective negative rate of change, sufficient to offset any effects of increased duration. Concerning adaptation for the stimuli in the patterned design, it has been suggested by Arntz *et al.* (1991) that changing levels of painful stimulation have a dishabituating effect. One problem with the adaptation type of explanation in our case concerns the relatively short durations in this experiment. It is true that some adaptation mechanisms work very fast, but it is not clear whether this is the case for our type of thermal pain. Hence, additional research is required to test the adaptation explanation, particularly for the difference between constant and patterned stimuli.

To summarize, four main findings emerged from the first experiment. First, the result showing that the final intensity is an important determinant of the overall retrospective evaluation was replicated. Second, it was shown that the intensity slope (rate of change) at the latter part of the experience had a positive impact on the retrospective evaluations. Third, it was shown that an increase in the rate of change leads to higher retrospective evaluations. Fourth, and perhaps most interesting, the duration of the painful experience impacts the retrospective evaluation for pain under some circumstances, but not others. Namely, prolonging the duration of the painful experience increases retrospective evaluations when the intensity levels of the painful experience change over time, but not when the intensity levels are constant. The main goal of Experiment 2 is to explore more fully the duration by Intensity-Pattern interaction that was observed in the first experiment, and to do so with the additional condition of on-line measures and within a different pain domain.

EXPERIMENT 2

Due to the limitations imposed by the Thermode equipment, duration in the first experiment was manipulated only to a small degree (10 seconds and 14 seconds). Therefore it seems appropriate to test the effects of pain duration more broadly. In particular, in order to look more directly at the question of pain duration and its interaction with Intensity-Pattern, one needs to sample a broader range of duration levels. Experiment 2 was designed specifically to test the combined effects of duration and level of intensity change (Intensity-Patterns) on retrospective pain evaluations. In addition, the issue of on-line measures and their possible intrusiveness was brought to bear in this experiment.

It is important to note that the results of Experiment 1, particularly with respect to duration effects, differ from those of Kahneman *et al.* The procedures used in Experiment 1 also differ from theirs in some very important ways. First, the current research used intensity levels that are above the pain threshold (unlike Kahneman *et al.*, 1993) and with a non-drugged population (unlike Redelmeier and Kahneman, 1993). In addition, the levels of pain intensity were physically manipulated, whereas the intensity levels in the Kahneman *et al.* studies were inferred from subjects' on-line evaluations. This latter difference has two separate consequences: one has to do with the stimulus Intensity-Pattern and the second with the subjects' responses.

In the Kahneman *et al.* studies, the stimulus intensity was relatively constant (sometimes with a minor change toward the end of the longer experience) and subjects were asked to continuously evaluate its intensity. In contrast to this method, the current work experimentally manipulated the intensity of the stimuli over a relatively wide range. In addition, subjects in Experiment 1 were not asked to continuously evaluate their pain intensity. Instead, I assumed that there is a monotonic relation between the perceived pain and the manipulated intensity. Perhaps this assumption is wrong. However, experimentally manipulating the intensity levels rather than using the on-line ratings has the additional advantage of avoiding any possible response contamination from the on-line evaluations to

 © 1998 John Wiley & Sons, Ltd.

the retrospective evaluations. Therefore, another purpose of Experiment 2 is to test whether on-line measurements coincide with the intended stimuli intensity and whether they contaminate the retrospective evaluations.

To summarize, Experiment 2 examines the effects of duration and intensity changes on the retrospective evaluations of constant and non-constant (patterned) stimuli. This is done in two response conditions: In the on-line present (On-line+) condition, where subjects give both on-line and retrospective evaluations, and in the on-line absent (On-line−) condition, where subjects give only retrospective evaluations (as in Experiment 1). In addition to extending and partially replicating Experiment 1, Experiment 2 also uses a different pain manipulation, namely pressure. The hope was to generalize the relationship between stimulus patterns, momentary response, and overall retrospective evaluations to a larger context of painful experiences.

In addition to the on-line measurements and the use of a different pain elicitation device, Experiment 2 differed from Experiment 1 in another important way. From the results of Experiment 1, it seemed that the effect of duration in the constant 45°C stimuli was due to the stimulus crossing the pain threshold with time. That is, for a few subjects, the shorter duration did not bring the experience above the pain threshold but the longer duration did. This problem did not occur with the other (more painful) stimuli since they were all clearly above this threshold level. In order to avoid this problem of crossing the pain threshold, one can use stimuli that are clearly above the pain threshold for all subjects. Alternatively, one can use stimuli that are personally calibrated to be minimally above the pain threshold for each subject. The use of such calibration technique seems desirable on two grounds. First, this method necessarily involves much less pain on the part of the subjects. Second, such a calibration method assures a higher subjective match between the experiences of different subjects.

The specific experimental hypotheses are:

- H5: On-line evaluations of the stimulus intensity map closely to the intensity manipulation. That is, the perceived intensity at each moment is highly related to the pressure manipulation.
- H6a: Global, retrospective evaluations are influenced by final intensity and final rate of intensity change (analogous to Experiment 1).
- H6b: Global, retrospective evaluations are influenced by the experience duration when there are changes in pain intensity over time, but much less so when there are no changes in pain intensity over time (analogous to Experiment 1).
- H7: The use of on-line measures has an intrusive impact on the retrospective evaluations. That is, when making on-line ratings the retrospective evaluations are based not only on the perception of the experience, but also on the produced pattern of momentary responses.

Method

Subjects

Forty subjects, mainly graduate students from the University of North Carolina and Duke University, participated voluntarily in a one-hour pain perception experiment. Subjects were 22 to 39 years old and all in good health. Sixteen of the subjects were female.

The equipment

A high-quality, precise, and stable vise (Brand and model: Record 2100B) was used to exert controlled pressure on the subject's selected finger. The calibration of the vise was tested on a large width range (82 mm) using a high-precision caliper. In this test, each complete turn of the vise's handle changed the

© 1998 John Wiley & Sons, Ltd.

width by precisely 4.1 mm. This linearity and sensitivity across the range assured both accuracy and safety for the subjects.

I should note that I personally used this procedure on myself and colleagues with pressure levels twice as large (2.4 mm) as used in the experiment, and with no lasting damage or pain. Such self-testing of equipment is the most common and safe method of developing procedures and techniques in this type of research. Once a safe range was established, a caution factor was taken such that subjects experienced no more than 50% of the pressure established as safe.

The stimuli

Each subject experienced two identical blocks, each composed of 21 stimuli presented in a random order. The Constant stimuli were: Constant Low, Constant Medium, and Constant High, and the Pattern stimuli were Up, Up&Down, Down, and Down&Up. Each of those seven stimuli were presented for 10, 20 and 40 seconds' duration, making the total of 21 stimuli in a block. The specification of the stimuli is similar to and based upon those used in Experiment 1. Each subject experienced one block in the On-line+ response condition and one in the On-line− response condition. The order of these blocks was counterbalanced between subjects.

Procedure

After being thoroughly instructed and familiarized with the equipment, each subject sat in front of a computer monitor with one finger placed gently in the vise. The vise itself was placed behind a curtain such that the subject could not see his or her finger, the vise, or the movements of the experimenter. This was done in order to eliminate any external cues regarding the pressure or the experimental manipulations. The finger was placed in the vise such that the cushion of the finger was placed against one side of the vise and the back part of the finger faced the other. The part of the finger that was in the vise extended from the middle point between the PIP and the DIP (the middle and last joints in the finger respectively) to the tip of the finger. At that point the experimenter closed the vise slowly until the subject indicated that he or she just began to feel its pressure as painful. This point, which is the minimal pressure level at which there is pain, is called the pain threshold, and it was used as the baseline for that subject. This point was marked on the scale of the vise and it was the starting point for each of the trials.

Next, in order for the subjects to become familiar with the procedure, the maximum pain intensity was presented by slowly decreasing the width of the vise. The maximum width decrease of the vise was 1.2 mm. Before increasing the pressure, subjects were told that the pressure would be slowly increased to its maximum in the experiment. In addition, subjects were told that the pressure of the vise would be terminated immediately at any indication from the subject. Once subjects were satisfied with the procedure, the experiment commenced.

Each subject participated in two experimental blocks, one in the On-line+ condition and the other in theOn-line− condition. In both cases, the subjects typed in their overall evaluation at the end of each trial by using the same scale as in experiment 1. In the On-line+ condition, subjects also tracked their moment-by-moment perception by continuously moving a cursor along a 0–100 scale. The cursor was moved on the computer's screen by using the right and left arrow keys on the keyboard. Each press on the right/left arrow key moved the cursor up/down 5 units (1/20 of the scale). The initial position of the cursor at each trial was always set at the middle of the scale (50).

Results

I analyzed the data with respect to four issues. The first involves the manipulation check (Hypothesis 5) and deals with whether the on-line response pattern followed the intended physical one. This

 © 1998 John Wiley & Sons, Ltd.

Exhibit 6. A conception of the different aspects of the data to be tested in Experiment 2

	On-line+ Condition (On-line & Retrospective)	On-line− Condition (only Retrospective)
On-line evaluations	A	
Physical pressure (manipulation)	C	D
Retrospective evaluation	E	F

hypothesis is represented as the relationship between the measures in cells A and C in Exhibit 6. The second issue concerns the effects of the different Intensity-Patterns on the retrospective evaluations, and more specifically their effect on the expected interaction between duration and the level of intensity change (Hypotheses 6a and 6b). These hypotheses are represented as the relationship between the measures in cells E and C, and cells F and D in Exhibit 6. The third issue concerns the intrusiveness of the on-line measure (Hypothesis 7), which is represented as the relationship between measures in cells E and F in Exhibit 6. The fourth issue is to model the relationship between the continuous aspects of the stimuli and the retrospective evaluations, seen as the relationship between the measures in cells A and E and cells A and F in Exhibit 6.

Analysis of the experimental manipulation

The first issue was whether the experimental manipulation was successful. In order to answer this question, the mean momentary intensities were plotted as a function of time. The three panels in Exhibit 7 present the data for all the Intensity-Patterns of 40 seconds' duration (the patterns were essentially identical for all three durations). As can be seen from Exhibit 7, the experimental manipulation was successful, in that the on-line response patterns essentially follow the intended intensity, thereby supporting Hypothesis 5. An additional issue is a methodological one concerning the first few seconds of the experience. As can be seen from the left side of each panel in Exhibit 7 (marked as the shaded area), subjects were a little delayed in reaching the intensity level that they wanted to report. This is so because the intensity of the stimuli themselves started at either a high, medium, or low level while the location of the cursor always started from the mid point of the scale (50).

This delay in adjustment can be seen for all Intensity-Patterns and for all durations. Note that the delay occurred even with the decreasing Intensity-Patterns, suggesting that this delay was not due to a

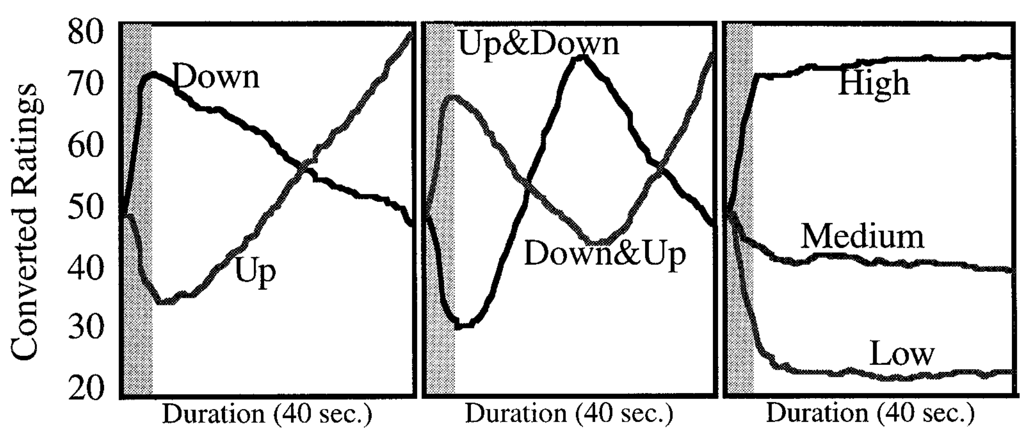

Exhibit 7. Traces of the On-line mean response for the different Intensity-Patterns in the 40-second Duration

© 1998 John Wiley & Sons, Ltd.

buildup of pain but rather to a delay in adjustment. In general, this result suggests that the experimental manipulation worked, but that there are potential difficulties in interpreting the initial On-line evaluations. Note that this problem is particularly crucial for the use of On-line evaluations with relatively short experiences. Since the assumption is that this discrepancy is between the reported and perceived intensities, the momentary responses for the first four seconds were eliminated from all remaining analyses.

Analysis of the experimental design

Overall, pain ratings collapsed across subjects had a mean of 53.2 and a standard deviation of 23.8, with a large between-subject variability in the use of the scale (see Exhibit 2). For the purpose of statistical analysis and in order to eliminate individual differences in the use of the pain scale, each subject's responses were again converted into T scores (see footnote 1), and all subsequent analyses were done on these scores. Using these measures, the data were first analyzed as a $7 \times 3 \times 2 \times 2$ mixed ANOVA design (Intensity-Patterns $\times$ Duration $\times$ Response-condition $\times$ Order of the two experimental blocks) with the first three factors as within-subject and the order condition as a between-subject factor. Block order was not statistically significant nor did it interact with any other factors and was therefore dropped from all further analyses. Looking at the experiment as a $7 \times 3 \times 2$ within-subject ANOVA design (Intensity-Patterns $\times$ Duration $\times$ Response-condition), the three-way interaction was not significant. All the two-way interactions were significant, Intensity-Pattern by Duration ($MSE = 83$, $F_{(12,468)} = 2.23$, $p = 0.001$), Response-condition by Duration ($MSE = 151$, $F_{(2,78)} = 5.82$, $p = 0.004$), Response-condition by Intensity-Pattern ($MSE = 391$, $F_{(6,234)} = 11.67$, $p < 0.001$). In addition, the main effect for Duration was significant ($MSE = 1004$, $F_{(2,78)} = 27.5$, $p < 0.001$), and so was the effect for Intensity-Pattern ($MSE = 16{,}697$, $F_{(6,234)} = 355.2$, $p < 0.001$). The main effect for Response-condition (On-line+ versus On-line−) was not significant. We now turn to look at these results in more detail.

Effects of Intensity-Pattern

First, it is interesting to note that the rankings as well as the ratings of the Intensity-Patterns were very similar to the ranking in Experiment 1 (see Exhibit 8). When testing the differences between the intensity ratings in Experiment 2, all but the difference between the High and Up Intensity-patterns were statistically significant (see Exhibit 9). The conclusion, therefore, is that once again the trend, as well as the final intensities, seems to be a determining factor in the overall retrospective evaluation of pain, thus supporting Hypothesis 6a.

Effects of Response-condition

Turning now to look at the simple contrasts in the Response-condition by Intensity-Pattern interaction, and examining the difference between the two response conditions for each Intensity-Pattern, all the differences but the ones for the Low and the Up&Down intensity patterns were statistically significant (see Exhibit 10).

It is interesting to look at the direction of those effects and note that for the constant stimuli as well as for the stimuli with final decreasing intensity, the evaluations in the On-line− condition were lower than the evaluations in the On-line+ condition. However, the results for the Up and Down&Up Intensity-Patterns were in the opposite direction. This finding indicates a dependency between the stimulus trend and the impact of the on-line measurements. This dependency suggests that in the case

 © 1998 John Wiley & Sons, Ltd.

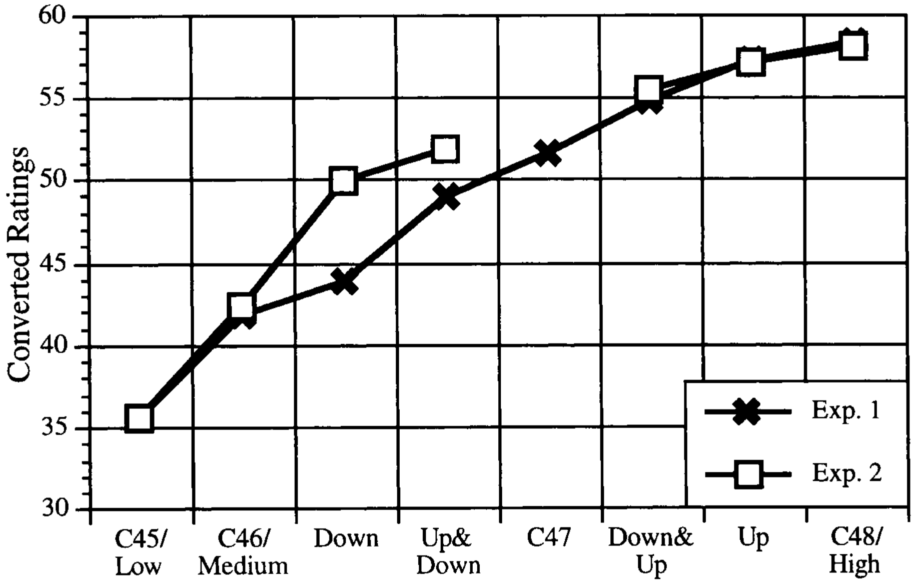

Exhibit 8. Mean converted scores for the Retrospective evaluations of the different Intensity-Patterns collapsed over the different Durations. The figure shows the data for all Intensity-Patterns in both experiments. The names of the stimuli refer to both experiments

of the On-line− condition there was a higher reliance on the pattern of experiences and that this reliance was reduced for the On-line+ condition. In other words, producing the continuous ratings in the On-line+ condition made subjects use this information in their retrospective evaluations. Therefore, the effect of the measurement condition had an opposite effect on stimuli with decreasing and increasing final trends, thus giving initial support for Hypothesis 7.

Exhibit 9. Contrast comparisons for all adjacent pairs of Intensity-Patterns in Experiment 2

Comparison	Difference	F-value	p-value
Low and Medium	6.77	117.1	<0.001
Medium and Down	7.59	147.0	<0.001
Down and Up&Down	2.00	10.2	0.002
Up&Down and Down&Up	3.57	32.6	<0.001
Down&Up and Up	1.63	6.8	0.009
Up and High	0.95	2.3	0.130

Exhibit 10. Effects of Response condition within the different Intensity-Patterns used in Experiment 2

Intensity-Patterns	On-line+	On-line−	Difference	p-value	Larger
Constant Low	36.03	34.96	+1.07	0.155	—
Constant Medium	43.88	40.66	+3.22	0.001	On-line+
Constant High	59.04	57.00	+2.04	0.007	On-line+
Up	55.24	58.89	−3.65	0.001	On-line−
Up&Down	51.99	51.73	+0.26	0.734	—
Down	51.40	48.32	+3.08	0.001	On-line+
Down&Up	54.55	56.32	−1.77	0.018	On-line−

© 1998 John Wiley & Sons, Ltd.

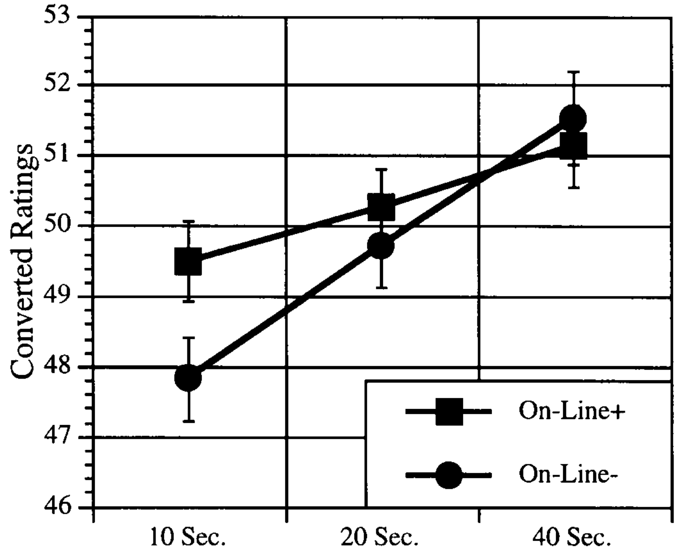

Exhibit 11. Mean converted scores for the Retrospective evaluations split by Duration and Response-Mode. Error bars are based on standard error

Duration in constant and patterned Intensity-Patterns

When looking at the main effect of Duration, all contrasts were statistically significant: the 10–20 second contrast ($\Delta\bar{X} = 1.33$, $F_{(1,78)} = 13.52$, $p < 0.001$), the 20–40 second contrast ($\Delta\bar{X} = 1.35$, $F_{(1,78)} = 13.99$, $p < 0.001$), and finally the 10–40-second contrast ($\Delta\bar{X} = 2.68$, $F_{(1,78)} = 55.01$, $p < 0.001$). These findings again support the idea that duration does matter. In addition, the interaction of the Response condition by Duration was also significant (as can be seen from Exhibit 11); the effect of this interaction was due to the fact that the impact of duration was smaller in the On-line+ condition compared with the On-line− condition. This difference, and the fact that the slope for the On-line+ condition was much flatter, is consistent with the idea that the retrospective evaluations in the On-line+ condition were affected to some extent by the on-line ratings and therefore were less sensitive to the passage of time, again giving more support to Hypothesis 7.

Turning now to the Intensity-Pattern by Duration interaction and testing the linear trend of Duration within each of the seven Intensity-Patterns, the results indicate that the trend in all but the Medium and Low Intensity-Patterns were statistically significant at the 0.01 level. This again indicates that for all but these two Intensity-Patterns, Duration did increase the perceived intensity. Now, collapsing within the constant and non-constant Intensity-Patterns and testing the contrasts for the different durations, the following results emerge: within the constant Intensity-Patterns, which include the Low, Medium and High Intensity patterns, neither the 10–20-second contrast nor the 20–40-second contrast were significant at the 0.01 level, but the 10–40-second contrast was ($\Delta\bar{X} = 2.36$, $F_{(1,468)} = 12.0$, $p < 0.001$). However, and as evident from Exhibit 12, the contrasts for the patterned Intensity-Patterns, which include the Down, Down&Up, Up, and Up&Down Intensity-Patterns, were all significant: For the 10–20-second contrast ($\Delta\bar{X} = 3.09$, $F_{(1,468)} = 29.7$, $p < 0.001$), for the 20–40-second contrast ($\Delta\bar{X} = 2.44$, $F_{(1,468)} = 18.59$, $p < 0.001$), and obviously significant for the 10–40-second contrast ($\Delta\bar{X} = 5.53$, $F_{(1,468)} = 95.24$, $p < 0.001$).

One of the main conclusions from Experiment 2, and in line with Hypothesis 6b, is that increased duration increases retrospective evaluations for patterned Intensity-Patterns, but less so for the

 © 1998 John Wiley & Sons, Ltd.

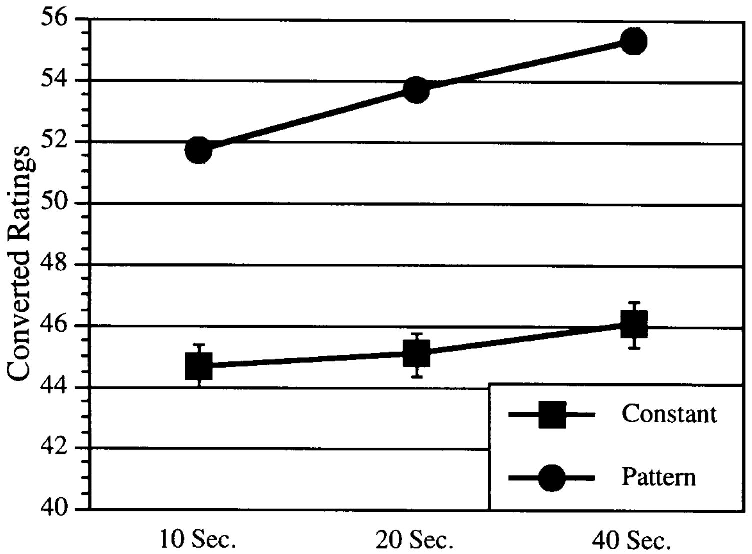

Exhibit 12. Mean converted scores for the Retrospective evaluations split by Duration and degree of intensity fluctuation in the Intensity-Patterns (constant versus patterned). Error bars are based on standard error

constant Intensity-Patterns (see Exhibit 12). Note that the High Intensity-Pattern was also affected by duration; however, this could be due to its extreme intensity (see on-line evaluation in Exhibit 7), therefore making adaptation impossible. To summarize, these results support hypotheses 6a and 6b and strengthen our belief in the findings of Experiment 1.

Intrusiveness of the on-line measure

As was evident from earlier sections, although Response-condition had no main effect, it did interact significantly with Duration as well as with Intensity-Pattern. The pattern of the interaction between Response-condition and Duration (as can be seen from Exhibit 11) suggested that the rate of increase in evaluated intensity was higher in the On-line− condition than in the On-line+ condition.

Regarding the smaller effect of Duration in the On-line+ condition, it is likely that the average on-line response produced by the subject was used in some way in the retrospective evaluation which in turn made the results less dependent on the duration of the experience. In addition, the finding regarding the difference between the Up and Down Intensity-Patterns suggests that subjects in the On-line+ condition were more affected by the mean response of the on-line evaluations. In other words, it seems that in the On-line+ condition, subjects used their own on-line responses in generating the retrospective evaluations. Another interesting observation is that the variances of the original responses (before they were converted to T) in those two conditions were statistically different, with the On-line+ condition having a smaller variance. The test on the variances was carried out on the natural log of the variance in order to meet the assumptions of normal distribution ($\Delta\bar{X} = 0.24$, $t_{(39)} = 3.2$, $p < 0.003$). This point regarding the variance could also be taken as additional evidence for a higher reliance on the subject's own produced responses. To summarize, evidence suggests that in the On-line+ condition subjects gave retrospective responses that were pulled toward their mean on-line ratings.

© 1998 John Wiley & Sons, Ltd.

Modeling the relationship between on-line and retrospective evaluations

In addition to analyzing the effects of the different experimental conditions on the retrospective evaluations, these data can also be used to test and compare different models that relate experienced intensity to the retrospective, global evaluation.

Model 1, which is also the most basic model, captures the assertion that subjects simply integrate the pain intensity over time. From the data in Experiment 1, and mainly from the analysis of pairs of patterns that were flipped around the middle of their vertical axes, Model 1 seems to be wrong. Nevertheless, it is the simplest and most straightforward model and testing it as such might provide a base from which to evaluate other models.

The scale used for the on-line responses had a total of 21 response categories, 0, 5, 10, ... 95, 100. Let i index the categories in sequential order of magnitude, such that $i = 0, 1, \dots 19, 20$. Therefore each category magnitude equals $5i$. Let T_i be the total number of half-seconds for which the subject gave a rating of $5i$. Then Model 1 is:

$$RE = C + \sum_{i=0}^{20} \beta_i T_i$$

where RE is the Retrospective Evaluation, and C is a constant used for scale transformations. Therefore, Model 1 reflects the 'area under the pain curve' without regard to any specific ordering of these reported intensities. Note that the estimation for the endogenous parameters (that is, the parameters describing the shape of the pattern) was taken from the on-line evaluations of the On-line+ condition and not from the specifications of the external stimulus. Modeling the relationship between the global and the momentary evaluations was done twice, once for the On-line+ condition and once for the On-line− condition. In the On-line+ condition, both the global and the momentary evaluations were based on the subjects' responses in the On-line+ condition. While for the On-line− condition, the global but not the momentary evaluations were based on the subjects' responses in the On-line− condition, the momentary evaluations in this condition were based on the subjects' responses in the On-line+ condition. Note that the retrospective evaluations in the On-line+ condition refer to the same experience on which the on-line measures were based, while the retrospective responses in the On-line− condition refer to the same external stimulus, but not necessarily the same perceptual or experienced stimulus. Therefore, the overall fit of the model is expected to be higher for the On-line+ condition than in the On-line− condition. Nevertheless, one can use the On-line− condition to draw conclusions regarding the relative importance of the different predictors.

Results for the On-line+ condition were significant with $R^2 = 0.46$ ($F_{(21,818)} = 32.7, p < 0.001$), and the results for the On-line− condition were lower, yet significant with $R^2 = 0.16$ ($F_{(21,818)} = 7.6$, $p < 0.001$). Some additional insight can be gained by looking at the standardized coefficients of Model 1 (see Exhibit 13). Note that these analyses were carried out on the converted T scores, which reduces the magnitude of R^2 dramatically.[2] The regression analysis used in this and all other models is such that the standardized coefficients reflect the unique contribution of the parameter, given that the other parameters are already in the model (Type III estimation procedure). As can be seen from Exhibit 13, the standardized coefficients for the On-line− condition are much smaller than the ones for the On-line+ condition. However, the trend in both cases is similar. That is, the lower intensities have negative coefficients, the middle intensities have coefficients around zero and the higher intensities have

[2] However, the analysis for the raw scores yield the same pattern of results in regard to the relative magnitude of the standardized coefficients.

 © 1998 John Wiley & Sons, Ltd.

Exhibit 13. Standardized coefficients for model 1 in the On-line+ and On-line− conditions. Coefficients that are significant are set in bold. Significance level used is 0.05 for the On-line+ condition and 0.1 for the On-line− condition

	On-line+ standardized coefficients	On-line+ t-values	On-line− standardized coefficients	On-line− t-values
Time at 0	−**0.264**	−5.83	−**0.141**	−2.25
Time at 5	−**0.245**	−4.65	−**0.167**	−2.29
Time at 10	−**0.265**	−5.29	−**0.128**	−1.85
Time at 15	−**0.151**	−3.78	−**0.090**	−1.64
Time at 20	−**0.205**	−4.55	−**0.123**	−1.97
Time at 25	−**0.111**	−2.39	−0.021	-0.34^{ns}
Time at 30	−**0.098**	−2.33	+0.028	$+0.48^{ns}$
Time at 35	−0.070	-1.95^{ns}	+0.019	$+0.38^{ns}$
Time at 40	−0.041	-0.94^{ns}	+0.023	$+0.37^{ns}$
Time at 45	−0.016	-0.41^{ns}	+0.076	$+1.38^{ns}$
Time at 50	−0.050	-1.31^{ns}	+0.065	$+1.23^{ns}$
Time at 55	+0.063	$+1.47^{ns}$	+**0.163**	+2.73
Time at 60	+0.129	+3.33	+**0.125**	+2.33
Time at 65	+0.103	+2.63	+**0.094**	+1.73
Time at 70	+0.139	+3.59	+**0.108**	+2.02
Time at 75	+0.189	+4.31	+**0.166**	+2.74
Time at 80	+0.192	+3.68	+**0.205**	+2.85
Time at 85	+0.241	+5.04	+0.105	$+1.59^{ns}$
Time at 90	+0.215	+4.21	+**0.120**	+1.69
Time at 95	+0.241	+3.64	+0.114	$+1.25^{ns}$
Time at 100	+0.173	+3.78	+**0.157**	+2.49

positive coefficients. Since the R^2 are moderately high, particularly for the On-line+ condition, this pattern indicates that the existence of lower intensities decreased the retrospective pain evaluation while higher intensities increased the retrospective pain evaluation.

This conclusion was sustained when a more restricted version of Model 1 (Model R1) was used, in which every three response categories were combined into one. In this case, results for the On-line+ condition results were significant with $R^2 = 0.45$ ($F_{(7,832)} = 97.7$, $p < 0.001$), and the results for the On-line− condition were lower yet significant with $R^2 = 0.16$, ($F_{(7,832)} = 22.3$, $p < 0.001$). Looking at the overall fit of Model 1 and Model R1 (R^2) for each of the Response-conditions, the overall fits were essentially identical. Turning to examine the standardized coefficients for Model R1, the conclusion again remained the same, with the coefficients for the lower intensities being negative, the coefficients for the middle intensity being non-significant and the coefficients for the high intensities being positive (see Exhibit 14). Another advantage of Model R1 is that it has the same number of parameters as Model 2, which will be presented next.

Model 2 is a different type of model, reflecting the conclusions of Experiment 1 as well as the conclusions from the different studies mentioned earlier (Kahneman *et al*., 1993; Redelmeier and Kahneman, 1993; Hsee *et al*., 1991; Loewenstein and Prelec, 1993). These conclusions are that the retrospective pain evaluation is based on the peak and the final intensities of the experience (Kahneman *et al*., 1993; Redelmeier and Kahneman, 1993), on the rate of change in the experience (Hsee *et al*., 1991), and finally on some improvement measures of the experience (Loewenstein and Prelec, 1993). These combined notions can be seen in Model 2:

$$RE = C + \beta_1 p + \beta_2 f + \beta_3 m + \beta_4 i + \beta_5 d + \beta_6 r_1 + \beta_7 r_2$$

© 1998 John Wiley & Sons, Ltd.

Exhibit 14. Standardized coefficients for a more restricted version of Model 1 (Model R1), in the On-line+ and On-line− conditions. Coefficients that are significant at the 0.05 level are set in bold

	On-line+ standardized coefficients	On-line+ *t*-values	On-line− standardized coefficients	On-line− *t*-values
Time at 0–10	−**0.259**	−10.31	−**0.147**	−4.24
Time at 15–25	−**0.163**	−7.75	−**0.083**	−2.91
Time at 30–40	−**0.070**	−3.59	+0.023	+0.85
Time at 45–55	−0.002	−0.11	+0.098	+3.69
Time at 60–70	+**0.129**	+7.24	+**0.118**	+4.79
Time at 75–85	+**0.208**	+10.23	+**0.157**	+5.60
Time at 90–100	+**0.208**	+9.10	+**0.130**	+4.12

where again *RE* is the retrospective pain evaluation, *C* is a constant used for scale transformations, and p, f, and m are the parameters for the peak, final, and the mean on-line evaluations respectively. Note that these latter three parameters are subjective and are derived from the on-line responses given in the On-line+ condition. The next parameter, called the improvement parameter (i), was introduced by Loewenstein and Prelec (1993) and represents the pattern of changes in the intensity over time. The improvement coefficient is calculated by taking the decumulative of the sequence, and comparing it to a decumulative with a constant change and the same sum. For example, a sequence of 40, 40, 40, 50, 60, 70 is converted by subtracting each number from the maximum number (70), making 70 into 0 and converting the series into 30, 30, 30, 20, 10, 0. The decumulative of this series is 120, 90, 60, 30, 10, 0 and the decumulative with the same sum (120) and a constant rate of change is 120, 100, 80, 60, 40, 20. The sum of the point-wise differences for this example is 110, which is the measure of the degree of improvement of the sequence (in this case improvement is negative, since it is a deteriorating rather than an improving sequence). The interesting aspect of this parameter is that it seems to capture some basic aspects of the Intensity-Pattern.

The final three parameters in Model 2 are duration (d), the rate of change in the first half of the experience (r_1), and the rate of change in the second half of the experience (r_2), representing subjective measures of the intensity. The rate of change was calculated separately for the first and second halves of the experience because the manipulation of the Intensity-Patterns was sometimes (Up&Down, Down&Up) different in the direction of change between the first and the second half of the experiences. The rate of change was coded with five levels of slope: high positive, medium positive, no slope, low negative and high negative. Note that since the experimental manipulation was so successful, there is no real difference between defining these parameters in terms of the pain manipulation or in terms of its subjective on-line reports.

The overall test of Model 2 showed a good fit for the On-line+ condition with $R^2 = 0.68$ ($F_{(7,832)} = 257.6$, $p < 0.001$). The results for the On-line− condition were lower, yet significant with $R^2 = 0.38$, ($F_{(7,832)} = 71.8$, $p < 0.001$). Note that this overall model fits better than both the full and the restricted versions of Model 1. Regardless of the overall fit of the model, the most interesting aspect is again the relative magnitude of the standardized coefficients (as can be seen in Exhibit 15).

The interesting observation from Model 2 and Exhibit 15 is that for both conditions, the final rate of change (r_2) was the most important predictor of the overall intensity evaluation, followed by the initial rate of change (r_1). In addition, note that for the On-line+ condition the mean on-line response is also relatively high, which strengthens the conclusion from Model 1. This finding indicates that subjects are in some way integrating the 'area under the pain curve'. However, since the mean-response is ranked as much less important (and negative) in the On-line− condition, one might speculate that the finding that subjects integrate the pain over time is an artifact of the on-line measure itself. Looking at the

 © 1998 John Wiley & Sons, Ltd.

Exhibit 15. Standardized coefficients for model 2 in the On-line+ and On-line− conditions. Coefficients that are significant at the 0.01 level are set in bold

	On-line+ standardized coefficients	On-line+ *t*-values	On-line− standardized coefficients	On-line− *t*-values
Peak	+**0.105**	+4.68	+**0.235**	+6.72
Final	+**0.112**	+5.17	+**0.134**	+3.96
Duration	+**0.044**	+2.51	+**0.079**	+2.88
Rate of change 1	+**0.598**	+2.49	+**1.060**	+2.82
Rate of change 2	+**0.726**	+2.54	+**1.576**	+3.52
Mean response	+**0.127**	+4.88	−**0.150**	−3.66
Improvement	−0.000	-0.64^{ns}	−0.000	$+0.89^{ns}$

other predictors in Model 2, it is clear that the Peak and Final intensities are important. Duration is also important but to a smaller degree. The only non-significant predictor in Model 2 is the improvement coefficient.

Discussion

Experiment 2 was carried out for three purposes. The first was to replicate the results of Experiment 1 and to do so within a different pain domain. The second purpose was to examine more carefully the effects of Duration and, in particular, to do so in the context of constant and patterned stimuli. The third purpose was to examine issues regarding the intrusiveness of on-line measures and their effects on the retrospective evaluations. The first analysis established that the experimental manipulation produced the desired experience pattern (Hypothesis 5). With this finding in mind, it was shown that the results of Experiment 1 were replicated with regard to the overall evaluations of the different Intensity-Patterns as well as with regard to the role of Duration in constant and patterned stimuli (Hypotheses 6a and 6b).

Finally, the data raised issues with regard to whether the demand to produce on-line measurements affects the retrospective evaluations. In particular, it appears that when producing retrospective evaluations, subjects show a dependency on their own on-line responses. It appears than that once subjects are asked to produce on-line responses, the saliency of such responses increases and hence they impact the retrospective evaluations (see also Dar and Leventhal, 1993; Schwarz and Bless, 1992). Evidence for this conclusion came from the differential impact of duration on the subjects in the two Response-conditions, and from the differential effect of Response-condition on the Up and Down Intensity-Patterns. More specifically, the Up Intensity-Pattern was rated less aversive in the On-line+ condition, and the Down Intensity-Pattern more aversive in the On-line+ condition. This finding supports the idea that in the On-line+ condition, subjects relied more heavily on their produced responses and, more specifically, were 'pulled' toward the average of these responses. Finally, the importance of the mean response in predicting the retrospective evaluation was higher in the On-line+ condition compared with the On-line− condition. It is important to note that not only did the importance of this predictor drop (as expected because of the reliance on different experiences), but it also changed its rank order from the third most important predictor in the On-line+ condition to having a negative relation in the On-line− condition. Again, this indicates that in the On-line+ condition subjects relied to some extent on their own momentary responses.

A final result of Experiment 2 relates to the pain experienced throughout the entire experience. At the end of the experiment subjects were asked to evaluate the overall amount of pain during the entire

© 1998 John Wiley & Sons, Ltd.

experiment (using the same scale as all their previous responses). The interesting aspect of this measure is that although it is a global retrospective measure, it summarizes multiple experiences and not just a single one. In a similar way to the approach of relating momentary and global evaluations, one could relate this end-of-experiment measure to the pattern of local experiences created by the whole experiment. While the current experiment does not allow one to examine such effects with much power (only one measure per subject) it is interesting to note that these end-of-experiment evaluations were only related to the mean response given by the subject and not to any aspect relating to the pattern of experience. Therefore, I would like to suggest that while the pattern of an experience has a large impact on its overall evaluation, a pattern of multiple experiences might not. This finding seems to have an intuitive appeal. For example, while the hedonic profile associated with a visit to a doctor might be a good predictor for our overall evaluation for that visit, it is hard to believe that the hedonic profile of that visit will merge with the hedonic profile for dinner that night (or, for that matter, with the rest of our life's experiences). It is therefore clear that the rules of experience summation for single and multiple episodes are fundamentally different.

GENERAL CONCLUSIONS

The two experiments presented here showed that characteristics of an experience's pattern determine much of its retrospective global evaluation. The aspects of the pattern considered important were the direction of change, the intensity of the slope (especially in the latter half of the stimulus), and the final intensity. Duration also had a role, albeit small; this role was larger for stimuli with fluctuating intensities than for more constant stimuli. Finally, on-line ratings were shown to moderate the retrospective evaluations. An interesting question remains: does the existence of on-line measures change the perception or memory of the painful experience, or does it only change its reported intensity in the experiment? If the latter is correct, then the effect is only an experimental artifact. However, one cannot yet dismiss the possibility that on-line measurement creates a fundamental change in the perception and memory of pain.

Although the existence of the on-line measure clearly had an effect on the retrospective evaluations, it was comforting to find that the order of blocks in the second experiment had no effect. That is, although some of the subjects reported that the On-line+ condition made them aware of the on-line ratings and that they kept on using this method even in the On-line− condition, the data showed no evidence for such carryover effects. The impact of the On-line+ condition was interpreted as an increased reliance on the responses subjects produced on-line. This reliance, whether deliberate or not, caused the retrospective evaluation in the On-line+ condition to be closer to the mean momentary pain.

An additional methodological issue concerning the on-line measures was raised, namely the discrepancy between the physical and reported intensities at the initial part of the experience. This observation implies that on-line measures should be used with caution with regard to the initial experience, and that the usefulness of this method for short duration is questionable.

Regarding the experimental manipulations in both experiments as well as the conclusion from Model 2, it seems that there are several aspects of the experience pattern that impact its overall evaluation. First and foremost, it is clear that the rate of change, and particularly the rate of change at the end of the experience, is an extremely important determinant of the perceived overall intensity. Evidence for this notion came from the difference between the different Intensity-Patterns and particularly from the difference between the Intensity-Patterns that were flipped around their middle vertical axis. More direct support for this idea came from the difference of the Time-Relationship in Experiment 1 and from Model 2.

 © 1998 John Wiley & Sons, Ltd.

The effect of duration was relatively small, yet in both the first and second experiments it was different for the constant and patterned stimuli. The impact of duration on the patterned stimuli, and its lack of impact on the constant stimuli, were interpreted as an adaptation-related phenomenon. The mechanism for this adaptation is not yet clear and more work is needed in this direction. Two possible adaptation mechanisms are the physiological and memory-based adaptations. The idea behind the physiological adaptation is that the longer an experience is, the less nerve activation it causes and therefore the less noticeable it becomes. The memory-related mechanism has to do with the relationship between perception of duration and the amount of change that occurred within that duration. It has been suggested that sensitivity to time depends to some extent on the amount of change within that duration (Avant, Lyman and Antes, 1975; Poynter and Holma, 1985). Therefore, when there is little intensity change (such as in the constant Intensity-Patterns), it is likely that there is decreased sensitivity to the intensities of the experience. Finally, the use of on-line measurements made the effect of duration even smaller.

The effects of the peak and final intensities were also strong. However, in order to examine these effects more deeply and to be able to separate them out, one needs to use many more Intensity-Patterns that are different on these parameters. The importance of the final intensity was supported by both experiments as well as by Model 2. In both experiments, it was shown that the retrospective evaluations for Intensity-Patterns that were flipped around their middle vertical axis were always much higher for the Intensity-Patterns that ended in high intensities compared with those that ended in low intensities (even though the overall amount of delivered intensity was the same). Additional support for this idea came from the importance of the final intensity parameter in Model 2.

Although subjects responded with no apparent difficulties to the request to produce momentary and global responses, one must wonder whether there are indeed two separate representations. More specifically, the issue is whether independent momentary representations exist at all. One strong argument against the existence of independent momentary evaluation stems from adaptation theory. From this perspective, there can be no pure representation of a current state (momentary evaluation) that is not influenced by the adaptation level. Therefore, it seems much more coherent to think of the momentary representation as one that includes the momentary stimulus as well as the history of the stimuli that preceded it. This, however, is very similar to our definition of the retrospective representation. Does this mean that the momentary and the retrospective evaluations are the same? The standard answer that more research is needed in order to understand this point seems very appropriate here as well.

Returning to the treatment of burn patients, it is clear that there is a large discrepancy between the momentary pain perception and its subsequent memory. An interesting question is which of these should the medical staff attempt to minimize? Is it the momentary experience or its memory? I would like to suggest that since future behaviors and choices are driven by overall evaluations, it is extremely desirable to decrease these memories. Imagine, for example, that a patient about to undergo a painful experience, his approach to the experience, his attempts to avoid it, and the stress and fears associated with the experience will be driven by memory of previous experiences. Hence the importance of reducing this global remembered intensity.

In this light of minimizing the overall remembered intensity, taking hold of the dressings (a short time with no pain) and then tearing them off in rapid short spurts (multiple short durations of intense pain with strong positive slopes) might not be the best strategy. This is because the integration of pain over time is not perfect, i.e. factors such as the final intensity and final rate of change have particularly high weight. The conclusion, therefore, is that bandages should be taken off slowly and steadily, which will cause a long duration for the treatment, but with a low intensity level, and with a low level of intensity change. Regarding the on-line measurements, it can also be suggested that allowing patients to continuously report their perceived pain intensity will increase their reliance on their reported pain and thus change their overall evaluation.

© 1998 John Wiley & Sons, Ltd.

ACKNOWLEDEMENTS

I would like to thank Tom Wallsten and George Loewenstein for their insightful comments and encouragement. I would also like to thank Frank Keefe and Jennifer Wideman for their help in the first and second experiments, respectively. Finally, I would like to thank my subjects for enduring in the name of science and friendship.

REFERENCES

Arntz, A., Van Eck, M. and de Jong, P. (1991). 'Avoidance of pain of unpredictable intensity', *Behavior Research and Therapy*, **29** (1991), 197–201.

Avant, L. L., Lyman, P. J. and Antes, J. R. 'Effects of stimulus familiarity upon judged visual duration', *Perception and Psychophysics*, **17** (1975), 253–62.

Cabanac, M. 'Physiological Role of Pleasure', *Science*, **173** (1971), 1103–7.

Carmon, Z. and Kahneman, D. 'The experienced utility of queuing: real time affect and retrospective evaluations of simulated queues', Working paper. Duke University, 1996.

Choiniere, M., Melzack, R., Girard, N., Rondeau, J., *et al*. 'Comparisons between patients' and nurses' assessment of pain and medication efficacy in severe burn injuries', *Pain*, **40** (1990), 143–52.

Dar, R. and Leventhal, H. (1993). 'Schematic processes in pain perception', *Cognitive Therapy and Research*, **17**(4) (1993), 341–57.

Dar, R., Ariely, D. and Frenk, H. 'The effect of past-injury on pain threshold and tolerance', *Pain*, **60** (1995), 189–93.

Fredrickson, B. L. and Kahneman, D. 'Duration neglect in retrospective evaluations of affective episodes', *Journal of Personality and Social Psychology*, **60** (1993), 45–55.

Hardy, J. D., Wolff, H. G. and Goodell, H. *Pain Sensation and Reactions*, Baltimore: Williams & Wilkins, 1952.

Helson, H. *Adaptation-Level Theory: An Experimental and Systematic Approach to Behavior*. New York: Harper and Row, 1965.

Hsee, C. K. and Abelson, R. P. 'Velocity relation: Satisfaction as a function of the first derivative of outcome over time', *Journal of Personality and Social Psychology*, **60** (1991), 341–7.

Hsee, C. K., Abelson, R. P. and Salovey, P. 'The relative weighting of position and velocity in satisfaction', *Psychological Science*, **2** (1991), 263–66.

Hunter, M., Philips, C. and Rachman, S. 'Memory for pain', *Pain*, **1** (1979), 35–46.

Jamison, R. N., Sbrocco, T. and Winston, C. V. P. 'The influence of physical and psychosocial factors on accuracy of memory for pain in chronic pain patients', *Pain*, **37** (1989), 289–94.

Kahneman, D., Fredrickson, B. L., Schreiber, C. A. and Redelmeier, D. A. 'When more pain is preferred to less: adding a better end', *Psychological Science*, **4**, (1993), 401–5.

Kahneman, D. and Tversky, A. (1979). 'Prospect theory: An analysis of decision under risk', *Econometrica*, **47** (1979), 263–91.

Kundera, M. *Immortality*, New York: Grove Press, 1991.

Loewenstein, G. F. and Prelec, D. 'Preferences for sequences of outcomes', *Psychological Review*, **100** (1993), 91–108.

Miller, N. and Campbell, D. 'Recency and primacy in persuasion as a function of the speeches and measurements', *Journal of Abnormal and Social Psychology*, **59** (1959), 1–9.

Perry, S. W. 'Under medication for pain on a burn unit', *General Hospital Psychiatry*, **6** (1984), 308–16.

Poynter, W. D. and Holma, D. 'Duration judgment and the experience of change', *Perception and Psychophysics*, **33** (1985), 548–60.

Price, D. D., McHaffie, J. G. and Stein, B. E. 'The psychophysical attributes of heat-induced pain and their relationships to neural mechanisms', *Journal of Cognitive Neuroscience*, **4** (1992), 1–14.

Price, K. P. and Tursky, B. 'The effect of varying stimulus parameters on judgments of nociceptive electrical stimulation', *Psychophysiology*, **12** (1975), 663–6.

Pritchard, R. M., Heron, W. and Hebb, D. O. 'Visual perception approached by the method of stabilized images', *Canadian Journal of Psychology*, **14** (1960), 67–77.

Redelmeier, D. A. and Kahneman, D. 'On the discrepancy between experiences and memories: Patients' real-time and retrospective evaluations of the pain of colonoscopy', Working paper, 1993.

 © 1998 John Wiley & Sons, Ltd.

Schwarz, N. and Bless, H. *Constructing Reality and its Alternatives: An inclusion/exclusion model of assimilation and contrast effects in social judgment*. Hillsdale, NJ: Erlbaum, 1992.
Stevenson, M. K., Kanfer, F. H. and Higgins, J. M. 'Effects of goal specificity and time cues on pain tolerance', *Cognitive Therapy and Research*, **8** (1984), 415–26.
Varey, C. and Kahneman, D. 'Experiences extended across time: evaluations of moments and episodes', *Journal of Behavioral Decision Making*, **5** (1992), 169–85.
Yarbus, A. L. *Eye Movements and Vision*, New York: Plenum, 1967.

Author's biography:
Dan Ariely considers himself a professional student. During his education he has attended Tel-Aviv University, the University of North Carolina at Chapel Hill, and is currently at the Fuqua School of Business at Duke University. His interests include decision making, consumer behavior, information integration, and pain as a hobby.

Author's address:
Dan Ariely, The Fuqua School of Business, Duke University, Box 90120, Durham, NC 27708-0120, USA.

© 1998 John Wiley & Sons, Ltd.