Summary Assessment of Experiences:
The Whole is different from the Sum of its Parts[1]

Dan Ariely and Ziv Carmon

Massachusetts Institute of Technology
INSEAD

Common experiences such as watching a film, waiting to be served at a restaurant, or undergoing a medical procedure, unfold over time through a stream of transient states that may vary from moment to moment in their intensity (e.g., become more or less pleasant) and even in sign (e.g., change from being pleasant to being unpleasant). A visit to the dentist for example, may begin with a boring wait in the reception room. Treatment can then begin with an unpleasant but not painful check-up, proceed with a painful drilling, and end with a lingering mildly painful sensation. Similarly, dinner at a restaurant can start with a fabulous appetizer, continue with an ordinary entrée, during which you become temporarily annoyed when you notice a piece of eggshell, which was clearly not supposed to be there, and conclude with a wonderful dessert.

Such evolution of subjective experiences can be depicted by an *experience profile*, whereby time is presented on the X-axis and the Y-axis represents the perceived momentary intensity of the experience. As an example, consider Figure 1 that presents the experience profile of a patient studied by Ariely and Carmon (2000). In that study, patients in the bone marrow transplant unit of a local hospital reported the pain they experienced on a 0-100 scale (0 represented no pain, and 100 the worst pain they could imagine) every hour from 8 AM to 6 PM. The figure indicates that the pain intensity for this patient was relatively high early in the day, then decreased, picked up around 1 PM, and generally improved thereafter.

••• Figure 1 here •••

Over the past decade a substantial body of research has examined summary evaluations of experiences such the ones described above. In particular, this research investigated the correspondence between experience profiles and their overall assessments, which is the topic of this paper. To illustrate, Ariely and Carmon examined how hourly pain reports such as those depicted in Figure 1, corresponded to end-of-the-day evaluations in which patients were asked to assess the overall pain they underwent throughout the entire day. A primary motivation for this line of research has been the finding that when people form summary assessments of experiences they *do not* combine the individual components of the experience profiles. Instead, a large number of studies has repeatedly demonstrated that neither the sum (integral) nor to the average of experience

profiles, corresponds closely to overall evaluations of their components (for reviews of this research see Ariely and Carmon, 2000, Fredrickson, 2000, Huber et al., 1997, Kahneman 2000). Therefore studying experience profiles must focus not only on their components (intensities of the transient states) but also on the rules people use to combine these components into overall evaluations.

Researchers have wanted to understand summary evaluations of experiences for several reasons: Overall evaluations of the pain and pleasure associated with different experiences are obviously important as input for future decision-making. How positively or negatively people remember an experience, is a key determinant of whether they will want to repeat it and whether they will recommend it to others. Retrospective summary assessments of events can also determine how people 'consume' memories of these experiences in the future. For instance, a brief exotic vacation can produce fond memories one could savor long after the experience is over. Like retrospective summary evaluations, prospective summary evaluations can also be important. They can evoke sensations such as anticipation and dread, before the experience ever takes place (Loewenstein 1987) and may thus determine whether or not one pursues an experience. Lottery tickets, for example, may often be purchased not so much because people truly expect to win the prize, but because they offer an opportunity to fantasize for a few days how it would feel to win the money, how to spend it, etc.

The rest of this paper is organized as follows. In Section 1 we sum up what is known about how experiences are summarized. In particular, we describe central features of experiences that appear to dominate their summary evaluation. We briefly describe empirical evidence for effects of these features and identify variables that can moderate those effects. In Section 2 we discuss the weighting of the duration of experiences in summary evaluations, and list variables that may influence this weighting. In Section 3, we present ideas about why the features we described in Section 1 affect summary evaluations. Specifically, we propose ideas about the role of efficient encoding, and about tendencies to predict future states. In Section 4 we describe ideas about retrospective reevaluations and reinterpretations of experiences, then suggest directions for future

research in this domain in Section 5. We conclude with a few final thoughts in Section 6.

1. Gestalt Characteristics of Experiences

An intuitively appealing way to summarize an experience would be to integrate, or perhaps average, across intensities of the subjective states of which the experience is composed (see Kahneman, 1994 and Kahneman, Wakker and Sarin, 1997 for detailed explanations of this argument). Yet, a clear conclusion of research in this domain is that when people summarize experiences, they do not integrate or average the transient states they experienced as the events unfolded. Rather, two types of defining features of the profiles of experiences appear to dominate overall retrospective evaluations. In Ariely and Carmon (2000) we name such defining features *Gestalt Characteristics*, of which there are two types— static and dynamic.

Static (state) *characteristics* reflect the intensity of the momentary experiences (i.e., transient states) at particular key points in time. To illustrate, a variety of studies have found that a weighted average of the momentary experience at the most intense and final moments, often referred to as the *peak* and *end*, respectively, can accurately predict the global retrospective evaluations (cf. Fredrickson and Kahneman 1993; Kahneman, Fredrickson, Schreiber and Redelmeier, 1993; Redelmeier and Kahneman, 1996; Varey and Kahneman, 1992).

Dynamic (configural) *characteristics* reflect the change in the intensity of the transient states as the experience progresses. Prominent examples of such characteristics include the trend of the profile (Ariely, 1998; Loewenstein and Prelec, 1993) and its rate of change (Hsee and Abelson, 1991; Hsee, Salovey and Abelson, 1994).

Note that it can often be difficult to distinguish between the two types of gestalt characteristics. For example, an experience with a very positive ending is likely to also have an improving trend, making it difficult to determine whether the ending or the trend made the overall experience as positive as it was.

Next we describe empirical evidence for the effects of these two types of gestalt characteristics.

1.1 Evidence for Effects of Static (State) Gestalt Characteristics:

The logic underlying Static Gestalt Characteristics is that people retain a few key statistics of their experiences, and it is those statistics rather than the complete experience, that are stored in memory for later use. An example of this notion of feature (statistic) retention in the context of assessment of experience profiles can be found in the work of Kahneman and Fredrickson (1993), who suggested that memory stores discrete snapshots of experiences. They offered a succinct metaphor inspired by Milan Kundera (1991), according to which "memory does not take film, it takes photographs." More generally, in a series of studies exploring different types of experiences, Kahneman and his colleagues repeatedly found that the differences between summary evaluations of different experiences is accounted for by a weighted average of the most extreme state (peak) and the final (end) state. In one study, for example, Kahneman and Fredrickson (1993) showed respondents either long or short movie clips, which were either pleasant (e.g., a puppy playing) or unpleasant (e.g., an amputation). Their respondents provided both continuous real-time ratings of their affective state, and a global retrospective rating of the pleasantness or unpleasantness of the experience after it was over[2]. They examined the correspondence between mean overall evaluations given at the end of each experience and the experience profile composed of the continuous on-line ratings respondents provided. Their results (as well as and those of others) indicate that people appear to rely only on key moments in their experiences when forming overall retrospective evaluations. In another interesting study, Redelmeier and Kahneman (1996) asked patients who underwent colonoscopy or lithotripsy to report the momentary pain they experienced during the procedure as well as to summarize the total pain they experienced after the procedure was over. Those results also suggested that a weighted average of the most intense moment and the final three minutes effectively predicted the retrospective summary evaluations of the experiences ($r = 0.67$).

1.2 Evidence for Effects of Dynamic (Configural) Gestalt Characteristics:

The logic underlying dynamic gestalt characteristics is similar to that we described for static ones. Individuals extract basic statistics of their experiences and it is these statistics, rather than the complete experience, that are stored for later use. But in the dynamic case the extracted statistics

represent *relationships* between different states of the experience rather than states in and of themselves. In particular, indicators of progress over time appear to be important to people (see Loewenstein 1987). Indeed, one of the most robust findings in research about assessment of experiences is the clear preference for improvement over time (see Loewenstein and Prelec 1993). Consider, for example, two different sequences of four dental treatments spaced over a week, for which the intensity of pain is represented by numerical ratings and larger numbers represent greater pain. In those two sequences the pain intensity either increases {2, 3, 4, 5} or decreases {5, 4, 3, 2} from one instance to the next. Although both sequences deliver the same cumulative discomfort, most people prefer the sequence of decreasing pain. To further illustrate, consider the example of the patients we described above (e.g., in Figure 1): In that study we found that ratings of overall daily pain were best predicted by the intensity of the final state and the slope of changes in pain ratings throughout the day (R^2 of 0.89). Moreover, in that case neither the average nor the sum of the pain experienced was effective at predicting the ratings of the overall pain.

In two other studies of dynamic gestalt characteristics, Ariely (1998) inflicted moderate levels of pain on participants either by using a heat probe or by pressing their finger in a vise. The experiences differed in their duration as well as in how they progressed over time (pain intensity over time either increased, or decreased, or increased and then decreased, etc.). After each experience, participants rated the overall pain they experienced, and those ratings were then regressed on features of the stimuli. The results suggest that participants were most sensitive to *changes* in intensity. Increasing pain intensity was perceived as very painful, and decreasing intensity was perceived as not painful, even when the sum of momentary intensities was the same. Also, some of the stimuli used in these experiments were not monotonic (i.e., increasing and then decreasing or decreasing and then increasing), and the results showed that sequences that first deteriorated and later improved were perceived as less painful than patterns that first improved and later deteriorated (see also Ariely and Loewenstein 2000). Thus, not only did people prefer improving sequences, they also preferred that the improvement take place later rather than earlier in the experience.

Preference for improvement has been demonstrated in many domains, including monetary payments (Loewenstein and Sicherman 1991, Langer, Sarin and Weber 2000), experiences such as vacations (Loewenstein and Prelec, 1991, 1993, Varey and Kahneman, 1992), queuing events (Carmon and Kahneman, 2000), pain (Ariely, 1998, Ariely and Carmon, 2000), discomfort (Kahneman, Fredrickson, Schreiber and Redelmeier, 1993; Ariely and Zauberman, 2000; Ariely and Loewenstein 2000), medical outcomes and treatments (Chapman, 2000; Redelmeier and Kahneman, 1996), gambling (Ross and Simonson, 1991), and academic performance (Hsee and Abelson, 1991, Hsee, Salovey and Abelson 1994).

1.3 Moderators of the Effects of Gestalt Characteristics:

The role of gestalt characteristics appears to be influenced by a variety of different factors. Examples include the type of experience, expectations, and the cohesiveness of the experience. These moderators are important because of their direct effect on overall evaluations, but also since they can provide insight into mechanisms that underlie people's summaries of their experiences.

One moderator of gestalt characteristics is the *type of the experience*. Carmon and Kahneman (2000) found that global retrospective evaluations of queuing experiences (experiences consisting of waiting) are dominated by the final affective state (end). In that setting, summary assessments effectively ignored the transient state components that preceded the ending of the experience. To illustrate, some queuing events that were dissatisfying up until seconds before they ended but concluded on a positive note, were summarized positively. To conceptualize the differences between the waiting experiences described by Carmon and Kahneman, and the more commonly studied experiences (Kahneman et al, 1993; Ariely 1998), we suggest that experiences fall on a continuum. On one end of the continuum are pure goal-directed experiences, such as fixing a flat tire, and waiting for a service. These experiences derive their meaning mostly from their outcome, after the experience is over. In such cases the momentary experience, as the event takes place, reflects perceived progress toward the goal whereas the summary assessment reflects the extent to which the goal is reached by the end of the experience (we discuss this idea further in Section 4). On the other end of the continuum are experiences that derive their meaning from the event itself

rather than its outcome. Examples include receiving a massage, fine dining, and watching a movie. Note that many real life experiences fall somewhere in between these two ends of the continuum. For example, activities such as driving to work and playing a game of squash are directed at goals of getting to work and winning, respectively. But at the same time, the on-going experiences in and of themselves are meaningful and they also influence the overall evaluations (how pleasant or unpleasant the drive was, and how good the game was).

Chapman (2000) proposes another moderator. She suggests that people's *expectations* moderate their desire for improvement over time for those events. For example, she found that subjects preferred that their skin initially appear young and become increasingly wrinkled with age, rather than preferring improvement over time as they do in most situations. On the other hand, she found that subjects preferred that another aspect of their facial appearance, the state of facial acne, improve over time. In another paper (Chapman 1996) she showed that people preferred declining sequences of health over time. Chapman argued that the preference for an escalating pattern is due to people's expectations regarding those experiences. For instance, she argues that since one's skin normally appears more rather than less wrinkled as one ages, people prefer the anticipated sequence of events. Indeed, in her studies Chapman finds a significant correlation between the patterns of events that people expect to encounter and the patterns they prefer. Read and Powell (forthcoming), however, suggest that the impact of expectations is often indirect, in that people view what they expect as appropriate and want that which is appropriate. When expected outcomes are not viewed as appropriate people prefer the appropriate.

Another moderator of gestalt characteristics is the *cohesiveness of the experience*. Ariely and Zauberman (2000) showed that the way people summarize an experience depends on whether they perceive it as composed of single or of multiple segments (i.e., if they see it as continuous or discrete). The authors show that the preference for improving trends over deteriorating ones is substantially reduced if the same experience is composed of discrete segments. Thus they demonstrate that the cohesiveness of an experience affects the relationship between its pattern and its overall evaluation. This seems reasonable in light of the idea that summary assessments partly

reflect inferences about future states. Specifically, if people naturally extrapolate current states in terms of their meaning for the future, the trend of an experience may appear more diagnostic of future states of continuous experiences than of future states of multiple discrete or segmented experiences (we will elaborate on this when we discuss naive extrapolation in Section 3.2).

Yet another moderating factor is the *'spread'* — the duration between the different segments of the overall experience. Loewenstein and Prelec (1993) showed that people preferred an escalating sequence when a later segment was to take place long after an earlier one, even though they displayed the more typical preference for improvement when the events were to take place within a relatively short period of time. For example, people would prefer to dine at an excellent restaurant before dining at a mediocre one, if the second dinner were to take place many months after the first. On the other hand, they would prefer the opposite pattern (first the mediocre dinner, then the excellent one) if the second meal were to take place a short time after the first. In other words, they show that the relationship between the pattern of the experience profile and its summary evaluation is moderated by the 'spread.'

In conclusion, the main finding reviewed in this section – that a weighted average of the static and dynamic gestalt characteristics of experiences predicts summary assessments – is robust across the many types of experiences. Variables we listed in this section (the extent to which the experience is goal directed, expectations regarding the experience, its cohesiveness and spread) moderate the relative weighting of these characteristics. Those three variables also moderate the desirability of improvement over time.

2. The Weighting of Duration in Summary Evaluations of Experiences

An intriguing implication of the role that different gestalt characteristics play in overall evaluations of experiences is the possibility that the duration of experiences does not significantly impact their summary evaluations (Kahneman, Wakker and Sarin, 1997). After all, if overall evaluations take into account only the gestalt characteristics (slope, peak and end) of the experience profile, what role remains for the duration of the experience? Indeed, an empirical generalization

that emerged from this research stream is that while other gestalt characteristics do influence overall evaluations, duration does not. Varey and Kahneman (1992) were first to draw attention to this phenomenon. In their studies subjects provided summary evaluations of each of several hypothetical experience profiles that differed in duration and in their intensity-pattern over time. Both Varey and Kahneman (1992), and Fredrickson and Kahneman (1993) found that summary evaluations of experiences corresponded closely to a weighted average of the maximum and final intensities of the experiences; while duration had almost no impact on the magnitude of the overall evaluations. Fredrickson and Kahneman (1993) termed this phenomenon "duration neglect," since after accounting for the gestalt characteristics of peak and end, duration had no significant effect.

In addition to documenting small marginal effect of duration beyond that of the maximum and final intensities, these authors also observed violations of dominance. For instance, subjects rated the overall pain in the hypothetical sequence {2, 5, 8} as worse than the overall pain in the sequence {2, 5, 8, 4}; where larger numbers represented more intense pain yet any positive number represented some pain. In other words, adding a segment with less severe pain, which represented a *relative* improvement at the end {4}, improved rather than hurt the overall evaluations of discomfort. Other examples of dominance violations were documented in a cold presser study (Kahneman et al., 1993), and for patients undergoing medical procedures (Redelmeier and Kahneman, 1996).

Although such results could be taken to indicate a complete neglect of duration, Ariely and Loewenstein (2000, see also Ariely, Kahneman and Loewenstein 2000) pointed out that all one can conclude from dominance violations is that subjects neither base their summary evaluations on the integral (sum) of pleasure or pain nor on a simple average (in which each is weighted equally). In fact, deviations from integration or simple averaging, such as giving special weight to peak, end, final slope, can produce violations of dominance, regardless of whether subjects do or do not attend to duration.

The original notion of duration neglect was recently modified to reflect the idea that rather than ignoring or underweighting duration, people do not evaluate sequences in the multiplicative

fashion predicted by discounted utility theory— a notion labeled "additive duration effect" (Schreiber and Kahneman 2000). Additive duration neglect implies that people do care about duration, but this does not depend on the intensity of the stimuli whose duration is varied. Discounted utility theory, in contrast, predicts that the impact of the duration of an experience depends on its intensity. For instance, people would presumably care more about how long a 110-volt shock lasts than about how long a 10-volt shock lasts. Additive duration effect implies that people's aversion to extending a shock would not depend on its intensity, which if true, could lead to undesirable decisions.

2.1 Effects of Response Mode on Duration Neglect

The phenomenon of duration neglect is intriguing. An important question both from a theoretical and a practical viewpoint is under what circumstances we may expect to observe it. In addressing this issue, Ariely and Loewenstein (2000, see also Ariely, Kahneman and Loewenstein 2000) point out two mechanisms that may influence the weighting of duration. The first is whether the goal of the judgment is to encode the overall goodness or badness of an event or to facilitate future choices among future events; And the second is whether the judgment is comparative or not (cf. Hsee et al. 1999). Below we describe those two notions.

Ariely and Loewenstein (2000) proposed that while the goal of choice is to maximize one's utility, the goals of encoding are more complex and typically include future use or communicating preferences to others. Under those goals, explicit neglect of duration can be sensible. For example, imagine that a friend asked you: "Overall, how would you rate your recent trip to San Francisco?" In that situation, the suitable answer would not incorporate the trip's duration, since you would mislead your friend if you were to rate a trip more positively simply because it lasted longer (except insofar as duration affected your average momentary pleasure from the visit). The typical reason to ask a question of this type is that the questioner is evaluating the desirability of visiting San Francisco for himself. The most useful answer, which would be in line with the questioner's expectations, would offer some average rating of your visit that does not take duration into account. If you responded "wonderful" because you had spent a full 2 weeks of slightly better than average

days in San Francisco, the questioner would be severely misled. The same would be true if you responded "awful" because you had spent only one, although fabulous, day in San Francisco. The questioner may not know how long you spent there, and you are unlikely to know in advance how long they might spend there if they were to go there. Indeed, the questioner may use your answer to the question, in part, to decide how long to spend in San Francisco.

The argument regarding situations in which you rate extended episodes as input into your own future decisions is similar. For a future decision of whether to repeat a past experience, it seems more useful to encode a summary measure of desirability that does not account for duration. This would allow the decision maker to decide about the duration of the future episode according to the conditions at the time of making that decision. If duration were encoded into the stored representation of desirability, and a decision maker were deciding whether to experience a new episode of different duration from the one he had experienced, the new judgment would require the decision maker to partial out the effect of duration from the evaluation recalled and then combine the new duration into it. Such an adjustment requires storing additional information (e.g., the duration of the original episode) and is, in practice, difficult to perform.

The second factor mentioned by Ariely and Loewenstein (2000) is whether the nature of the judgment is comparative or not (for extensive discussion of difficulties in judging attributes in isolation see Hsee et al., 1999, Nowlis and Simonson, 1997). Duration may be an attribute that is difficult to judge in isolation, without direct comparison to other events. In such situations Norm Theory (Kahneman and Miller, 1986) suggests that each evaluation will automatically evoke a norm of comparison– even if the judgment is not explicitly comparative. In rating a visit to the Grand Canyon, for example, people are likely to compare it to some other long trips they took. People are less likely to compare their San Francisco trip to a dinner or a sports event. The same notions apply to duration. When evaluating a particular morning's commute, people are unlikely to evaluate it relative to a recent cross-country drive. Duration is one of many variables that people use to classify stimuli for purposes of scale-norming. If people indeed norm on the duration of an experience (comparing short experiences to other short experiences and long experiences to other

long experiences), then observing a neglect of duration will follow by design.

To examine these issues Ariely and Loewenstein (2000) conducted a study using four different elicitation methods that varied on whether they involved ratings or choices, and whether evaluations were comparative or separate (see Table 1). As Figure 2 shows, results suggest that both factors are important to the weight respondents placed on duration. The role of duration appears to increase when responses are comparative rather than separate, and also when the responses involve choices rather than ratings. Additional evidence for the effect of separate vs. comparative judgments comes from recent work by Sonnenschein, and Shizgal (2001). In their experiments rats working for brain stimulation placed much greater weight on the duration of stimulation when the length of the reward could be compared to other rewards. When rewards were presented separately, the role of their duration was vastly diminished. Note that the role of other gestalt characteristics, such as the slope, final intensity, and peak intensity did not differ across the four different elicitation modes, suggesting that the role of duration is unique both in its sensitivity to different response modes and with regard to how it is used.

••• Table 1 here •••

••• Figure 2 here •••

2.3 Is Underweighing Duration an Error?

The relatively low impact of duration on overall evaluations (see Figure 2) may suggest that people under-weigh duration. But this assumes a normative model of how duration should be integrated, and it is not clear what that normative model would be. One complication is that people derive utility not only from the experience in and of itself, but also from anticipation and memories of the experience (Elster and Loewenstein, 1992). Because these sources of utility can be significant, it could be perfectly reasonable to prefer a longer sequence of pain (with a larger integral of utility), which leads to less disutility from memory or anticipation. To the extent that pleasure or pain from memory and anticipation are not themselves influenced by duration, it can be normatively defensible to weigh duration less heavily in choice. Thus, it can make good sense to

prefer a longer colonoscopy that ends on a good note to a shorter one that ends in excruciating pain if the longer procedure is remembered more favorably, or if the next one is dreaded less, even if the sum of discomfort during the longer procedure is greater.

Even if one could measure utility from memory and anticipation, however, which would be exceedingly difficult, it would still be questionable whether utility integration is a compelling normative principle. For many normative rules of choice, such as dominance (if A is better than B on all dimensions, then choose A) or transitivity (if A is preferred to B and B is preferred to C, then A should be preferred to C), many people are persuaded that the rule should be followed after it is explained to them, and they generally want to change their behavior if they are made aware that they violated the rule. This is not the case for utility integration. People often deviate dramatically from utility integration in prospective studies of preferences for sequences (e.g., Loewenstein and Prelec, 1993) and they do not tend to change their minds, even when the logic of doing so is explained (Loewenstein and Sicherman, 1991). People do care about properties of sequences other than the integral of utility that they provide, and the fact that they do so knowingly and unapologetically, should make us wary of labeling their preference a bias. We propose that future work explore in depth how and when duration is (and should be) integrated into overall judgments, both from descriptive and prescriptive perspectives.

2.4 Conditions Influencing the Weighting of Duration

We offer the following two distinctions to help understand effects of duration on summary evaluation of experiences. First, when attention is drawn to the duration of an experience, its role clearly increases. This can occur when the duration of the experience is an important characteristic (e.g., in experiences such as childbirth, prison sentences, or a wait in a restaurant; cf. Carmon et al. 1995), or when there is an explicit comparison across experiences such as comparing a 20-minute massage to a 40-minute one. A second factor is the extent to which duration is inherent to (i.e., an integral aspect of) the experience. To illustrate, the duration of an afternoon walk is fairly extrinsic to the experience, and a person can thus decide at any point during the walk about its desired duration. This is not as true of an experience such as a movie or a cruise, as the duration of such

experiences is more inherent.

The extent to which people will weigh duration depends on a combination of these two factors. We suggest that people will weigh duration least when attention is not drawn to this attribute and when the duration of the experience is inherent to the experiences. In other cases duration will be weighed more heavily. The role of duration in real decisions therefore depends on what form these factors take in day-to-day decisions (ecological validity).

3. Toward an Understanding of why the Gestalt-Characteristics Matter

The significance of the defining features (gestalt characteristics) of experiences in predicting summary assessments is by now widely accepted. A remaining research challenge is to understand why, how, and when these gestalt characteristics are important. We propose that reliance on gestalt characteristics has two main functions: efficiency in memory and ability to predict future states. Below we describe each of these in turn.

3.1 Memory Efficiency

One role of summary evaluations may be to cope with people's cognitive limitations. This requires efficient representation of the many detailed characteristics of stimuli such as experienced events. Indeed, we believe that the notion that people retain key features may be true not only for the assessment of experiences. To illustrate, in a recent study Ariely (2001) showed respondents sets of circles of dissimilar sizes for a brief duration of 500ms, followed by a single circle also displayed for a period of 500ms. On some trials participants were asked to indicate whether they had seen a circle of particular size within the set of circles they were previously shown (i.e., if a circle of that size was "a member of the set"). On other trials participants were asked to indicate whether the single circle was larger or smaller than the mean of the "set" of circles shown in the previous exposure. The results showed that as the number of circles in a set increased, recognition of individual circles (i.e., correctly answering questions as to whether the particular circle had been a member of the set of circles shown) dropped rapidly to chance level, but recognition of the mean of the set remained precise. Moreover, respondents appeared to have a good sense of the distribution

(variance) of the sizes.

We find it interesting that even with simple visual stimuli, people seem to extract and retain a few key statistics rather than the original complete set of information. Note that in order to represent information efficiently the visual system "could have" either represented all the information at low resolution (in less detail) or represented highly accurate statistical properties of the stimuli. That the visual system developed so as to accurately represent statistical properties suggests to us that there is an advantage to statistical representation, perhaps also in domains other than visual perception.

The application of these ideas to gestalt characteristics of assessed experiences is straightforward. For many situations parsimonious representation of experiences with a few key characteristics seems reasonable and highly adaptive. The alternative, representing and retaining each of the individual transient states of an experience, may demand too many cognitive resources, and offer only marginal benefits. Thus, representation by statistical properties seems efficient and effective. The focus on the specific statistics of the peak and the end seems adaptive because the peak is the aspect of the experience that may often signal the extent to which the experience may be risky, and the ending can often seem important for learning about the effectiveness of the course of action (for more on this, see the next section).

3.2 Predicting Future States

Another important goal of assessing and summarizing experiences is to facilitate effective decisions by helping to predict future states. To illustrate, imagine a patient undergoing a painful and long medical treatment that becomes less painful over time. Based on aspects such as a trend of decreasing pain, the patient may extrapolate that the future is likely to be less painful, or even infer that she is closer to recovery. This extrapolation-based explanation can help us understand why dynamic (configural) aspects of the experience (i.e., the manner in which its intensity evolved, such as trend) and the final state (end) are emphasized in summary assessments. The key is that these gestalt characteristics may help, or seem to help, predict future states.

We propose "naive extrapolation" as an underlying concept that can explain the strong

preference for improving sequences (see Loewenstein 1987 for related ideas). The notion is that decision makers naturally incorporate into their evaluations their expectations for the future, which they draw from how experiences progress over time. Hence, decision makers extrapolate the progression of experiences over time to predict their future state (even if the experience terminates). The notion of naïve extrapolation is based on the ideas that decision makers assess implications of the present state for the future, and that the anticipation of future states is incorporated into current evaluations. For instance, in the area of pain, decision makers who experience an increasing pattern of pain: (1) are likely to predict that the pain will continue to escalate (or remain high); and (2) this prediction, because of its anticipated negativity, is likely to make the present experience feel worse. Note that this incorporation of future states would be reasonable if the task indeed called for such predictions. However, we believe that decision makers do so even when there is no apparent reason to do so, and even when they are clearly asked for retrospective evaluations – explicitly instructed to ignore the future and evaluate only the past. We refer to this as the *naive extrapolation hypothesis*.

Currently, the evidence for the naive extrapolation view is very limited. Two experiments by Ariely and Zauberman (2002) are consistent with this idea. In one experiment, the authors showed that breaking an experience into segments reduces the effect of the trend (improvement vs. deterioration). The argument is that if people extrapolate from a trend, partitioning will decrease this tendency and thus reduce the impact of trend on summary evaluation. In a second more direct experimental test, the authors asked respondents not to provide overall evaluations but instead to predict the future state of the hedonic profile. The results showed that manipulations that decreased participants' tendency to extrapolate (partitioning, and increased unpredictability) also decreased the effects of the trend of the experiences on overall evaluations. The tentative conclusion is that at least to some degree, people automatically extrapolate the meaning of their current experiences to future ones and that these extrapolations influence how they experience the present.

4. Why the Past may seem Different than when it was Present

An interesting aspect of summary assessments of experiences that is not well understood is the drastic change in how people view some experiences retrospectively compared to how they felt

as the events took place. Rosy retrospective sentiments about one's military service, a stormy former romantic relationship, or what the aging refer to as 'the good old days,' to name a few examples, may well be vastly distorted. Those "misguided" feelings are often supported by specific memories that are consistent with the sentiment, but distort what was actually experienced. The memories may be a result of selective recollection and suppression of true events, significant misrepresentation of other events, and possibly even 'mental construction' of events that never actually happened but eventually seem very real nevertheless.

Loewenstein (1999) offers several examples of how differently people sometimes think of and remember an experience once it is over compared to how they felt about it as it was taking place. In an illuminating essay about the utility people derive from mountaineering experiences, he cites descriptions of conditions people undergo in this activity as "… harshly uncomfortable, miserable and exhausting..." To explain why people nevertheless engage in such an activity he further cites descriptions of the change in perspective between the way the activity is experienced as it occurs and how it is remembered after it is over. For example, Simpson, a renowned climber, reports that his perspective changed almost instantly after he reached the top of a mountain: "On the summit my memory edited out the anxiety and tension and fed me happy recollections of superb climbing." Stroud, a famous arctic explorer, offers similar observations: "Even though I can clearly remember saying to myself every day of the journey: I must never do this again' I don't feel now as I did then. The memory deficit is playing its tricks already."

Loewenstein's essay suggests that such experiences may be undesirable as they take place, but have desirable value (perhaps symbolic) after they end. He further lists a variety of reasons for 'consuming' an experience other than for its own sake or for the sake of its direct outcomes. The notion is that in some instances the ending of an experience gives it special meaning. Therefore, how people feel as the experiences take place may have little to do with how they assess and perhaps even remember the same experiences in retrospect. One such symbolic goal could be self-signaling (see Bodner and Prelec 2001; Dhar and Wertenbroch 2002) whereby a difficult activity can 'reveal character.' Loewenstein suggests that "This desire for a harsh test … along with poor memory for

misery, may help explain why the most miserable trips often produce the best memories; pain and discomfort are, to some degree, the *point* of the trip." Completion is another goal that could be related to self-signaling, because the completion of a task defines the task as a whole and thus the value of self-signaling rises mostly from completing a task and not from simply engaging in it. Mastery is yet another goal, about which Loewenstein says "It is generally pleasant to engage in an activity that you are good at, no matter how useless it might be… it is typically aversive to do something you are incompetent at, no matter how instrumental the activity." Finally Loewenstein also lists the search for meaning of life as a goal whose pursuit could undermine momentary utility and memory.

In a different approach to the same general topic, Tykocinski (2001) suggests that there are instances in which people may alter how they assess an experience after it is over in order to maintain how they feel about themselves (rather than in order to enhance it, as in Loewenstein's examples). Specifically, she observes that "…pointing out that a tragic event was inescapable, or somehow 'bound to happen' appears to be a popular tool in our solace repertoire." She proposes that to make disappointing events more palatable people sometimes alter perceived probabilities of relevant events after the fact. The underlying goal is to help deal with the unpleasant event by making it appear almost inevitable as well as making more positive outcomes seem highly unlikely. This idea offers another possible reason for substantial discrepancies between perceptions of experiences before, during and after they take place — self-deception.

Such motivated biases in memory can be different for different types of experience profiles, and different levels of gestalt characteristics. To the extent that some of the gestalt characteristics are more memorable than others, it is likely that memory distortions will occur more frequently and rapidly (and with less mental effort) for those aspects of the experience profile that are not remembered well. Such less remembered aspects could be the parts of the experience that are not captured by the gestalt characteristics, or by gestalt characteristics that are not very salient (a low peak, a low level of improvement etc.).

5. Directions for Future Research

In this section we highlight a few directions for future research that we view as important for better understanding assessments of experiences. We then end this chapter with a few concluding thoughts.

5.1 Ending Effects:

As described in Section 4, people sometimes view experiences very differently after the fact compared to how they felt as those experiences took place. We propose that future research explore when, why and how such retrospective 'reinterpretation' of experiences tends to happen. One specific approach to this would be to empirically explore ideas described in Loewenstein (1999) about effects of goals people might pursue as they undertake an activity. One could test, for instance, whether the extent to which an experience has a symbolic value and the degree to which the final goal was achieved influences the ways in which decision makers combine the experience profile into overall evaluations.

The effects of endings may be important not only when goal-attainment is defined by the final part of an experience, but also because endings define the boundary of the experience. Moreover, ending points are likely to be natural points in which decision makers summarize their experiences (Ariely and Zauberman 2000). A change in motivation toward the end has been noted in rats and pigeons who increase their effort as a function of temporal distance from the end of the experience – a pattern known as scalloping – even when the increased behavior does not improve their payoffs (Ferster and Skinner, 1957). In humans, the effects of endings have been reported to influence many different behaviors including impatience (Ceci & Bronfenbrenner, 1985), the propensity to get agreement in negotiation (Roth, Murnighan and Schoumaker, 1988), as well as bidding behaviors in online auctions (Roth, and Ockenfels, forthcoming).

To test such effects of endings, consider the following possible experiment (cf. Fredrickson 1991): subjects evaluating an experience would occasionally be prompted to indicate intermediate summary evaluations (a summary of their experience up until that point in time). At some point a subset of the subjects would be told that they had reached the end of the experience and asked to

provide the final summary assessment, whereas others would be lead to believe that they were merely providing another intermediate summary evaluation. Comparing evaluations across such a manipulation may offer insights into how assessments change when the end is reached.

The effects of endings can be important not only in retrospect but also prospectively. We find it interesting, for example, that patients are often very keen to know when their pain will end or how long treatment will last, even though such knowledge would not alter the actual events they will experience. What value such information may offer and how advance knowledge about the end can reduce the pain are interesting and important questions. Possible underlying factors include perceived control, allocation of coping resources, and decreased tendency to naively extrapolate (see Section 3.2, and Langer 1975 for related ideas). Support for such notions is apparent even from casual observations of people who exercise. Such people routinely want to know how they are doing in relation to their goals, in order to assess how much effort is required to accomplish these goals. This knowledge can presumably help sustain motivation and increase the probability of achieving the goal.

In a small test of this notion, 60 people exercising in a gym were asked to stretch their dominant arm to the side of their body (parallel to the ground) holding a 5 lb. weight for as long as they could. The duration for each participant *I* was recorded and termed X_i. Once the participant could no longer hold the weight, he or she was immediately asked to repeat the process using the other (non-dominant) arm. This time, however, the task was to hold the weight for a duration of X_i +30 (thirty seconds longer than they had been able to hold with their stronger arm). Participants were randomly assigned to one of three conditions. In the "up-counting" condition, the experimenter counted each passing second aloud (from 1 until the time goal, X+30, or until the participant stopped of his or her own accord). In the "down-counting" condition, the experimenter counted the seconds aloud (from X+30 down to 0, or until the participant stopped of his or her own accord). In the "no-counting" condition, the experimenter provided a free association every second, effectively preventing the participants from knowing how they were doing in relation to their goal.

Results show that the different counting conditions greatly affected how long participants were able to hold the weight with their non-dominant arm [$F(2,57)=43.48, p<0.001$; all the pair-wise differences $p<.001$]. Participants in the down-counting condition managed to hold the weight for the longest time (24.7 seconds longer than with their dominant arm, which was 5.3 seconds less than the requested goal). Participants in the up-counting condition managed to hold the weight for a shorter time (5 seconds more than with their dominant arm, which was 25 seconds less than the requested goal). Participants in the no-counting condition managed to hold the weight for less than all participants in the other two conditions; In fact, whereas those in the other conditions were able to hold the weight longer with their non-dominant arm, this was not true of participants in this condition(22.6 seconds less than with their dominant arm, which was 52.6 seconds less than the requested goal). Note that across the two conditions only two participants reached the goal, and the results hold if these two respondents are eliminated from the analysis- indicating that it is not simply having a known end point that causes people to reach their goal. In sum, these results show that knowing the end (comparing the up-counting and down-counting conditions versus the no-counting condition) increases tolerance. They also show that having a clearer view of an end point (the 0 in the down-counting condition compared with the X+30 in the up-counting condition) further increases tolerance .

Several questions arise from these findings. Why does knowledge of the end point improve coping ability and motivation? What tools can we provide patients to better cope with pain and thereby diminish it? Does the improved coping ability (cased by knowing when the end will arrive) hold for all time frames— is this knowledge as beneficial for a durations of months, weeks, hours, and minutes? Moreover, once the end point has been identified, do the benefits of knowing it increase as time goes by and the end draws nearer? Clearly, more research on the psychological effects of "end-knowledge" regarding endurance and coping ability is needed.

5.2 Memory Effects:

A research topic related to the effects of endings is how summary assessments and memories that are associated with them change as a function of the temporal distance from the event — soon

after the experience is over, some time afterwards, and long after that. For example, would the memory decay be the same for the different gestalt characteristics, or would some gestalt characteristics decay and be forgotten faster than others?

The effects of time may not be limited to memory decay. As suggested by the work of Loewenstein (1999) and Tykocinski (2001) memories of experiences may be constructed in a manner that ignores significant aspects of the original experience, thereby associating different meanings with the experience. It is thus important to go beyond merely examining the accuracy of overall evaluations of such experiences, and also assess whether the discrepancies are a result of distortions of specific aspects of the experiences.

Finally, memory can also change the grouping of different experiences into experience profiles. A simple example of this is that shortly after a vacation, each day of which it was composed may be perceived as an independent experience, while a few years later the boundaries of the individual days may well be blurry and the vacation is likely to be perceived as a single experience. Such grouping effects are not limited to blurry boundaries and can also occur as the meaning of experiences change. For example the integration of a single date with the rest of one's relationship with a romantic partner might depend on the length, success, and variability of the relationship.

5.3 The Evaluated Dimensions:

Another potentially interesting direction for future research is to examine different dimensions of people's evaluations of experiences. Research to date has mostly focused on a single measure of affect—the extent to which one feels positively or negatively about the experience. Yet as discussed in Section 4, it is clear that people's motives for pursuing an experience sometimes have little to do with how well it makes them feel as it takes place. In such cases it may be more productive to examine real-time momentary assessments that relate more directly to the goal the person is pursuing in that instance. The correspondence between the summary assessment of the experience and real-time experience profiles on those dimensions (rather than the affective experience profile) may be instructive. If you consider Loewenstein's mountaineering examples,

for instance, measuring the experience profile on dimensions such as mastery or goal completion may be more informative than how positively or negatively the person feels as the experience progresses.

Another potential avenue for future investigations is based on research that has identified common co-existence of several distinct types of emotions (it has been suggested for example that pain and pleasure are distinct emotions that do not define ends of one continuum). There is no reason to believe that those independent types of emotions would necessarily evolve in the same way throughout an event, or even be summarized in the same fashion. In some situations summary assessments of an experience may therefore be better predicted based on some weighted average of experience profiles along different emotions. In conclusion, while an overall measure of emotions, such as that used in most research to date, may often capture much of the needed information effectively, there are likely to be exceptional cases, such as experiences that evoke mixed (conflicting) emotions.

5.4 The Types of Experiences:

Much of the research that explicitly measured experience profiles and examined how they correspond to their summary assessments has focused on unpleasant events. Examples of the types of experiences that have been investigated include discomfort (Kahneman et al. 1993; Ariely and Zauberman, 2000; Schreiber and Kahneman, 2000; Ariely and Loewenstein 2000), medical procedures (Ariely and Carmon, 2000; Redelmeier and Kahneman, 1996; Katz, Redelmeier, and Kahneman, 1997), queuing experiences (Carmon and Kahneman, 2000) and pain (Ariely, 1998). Pleasant experiences, on the other hand, have received little attention. One reason that pleasant experiences have hardly been investigated is that it is difficult to cause pleasure to people in a controlled manner. Food and sex, for example, two major causes of pleasure, are not monotonic with stimulation intensity. That is, more of these two experiences is not always better.

Little attention has also been devoted to mixed experiences, that are at times pleasant and at other times unpleasant (cf. Kahneman 1992). Experiences can also be mixed in other ways. Largely unrelated events may co-occur, for example. Thus, one may experience pain in one part of

the body and pleasure in another. Or one can watch a good movie in the cinema while feeling the urge to urinate. It is not at all obvious how people would summarize such mixtures of events. More generally, investigating how people summarize different types and combinations of experiences may help identify additional moderating variables and border conditions. Such research is likely to provide further insights about summary assessments of experiences.

6. Conclusion

Roughly a decade of extensive research activity on the relationship between how events are experienced as they occur and how they are retrospectively summarized has resulted in a considerable body of knowledge. In this paper we offered a critical overview of selected aspects of this literature. We summarized a handful of findings that are widely accepted. In particular, we described central features of experiences (gestalt characteristics) that seem to govern summary evaluations of those events, and identified variables that can moderate those effects. We next discussed the weighting of the duration of experiences in summary evaluations, and listed variables that could affect this weighting. We continued with a discussion of why the gestalt characteristics we described early in the chapter affect summary evaluations, and proposed ideas about why some experiences are remembered very differently from how they were experienced as they occurred. We concluded by presenting directions for potential future research.

A more complete theory of summary assessments of experiences could offer a better understanding of people's preferences in a very broad variety of domains. Beyond obvious theoretical value, this could have important practical implications, since many types of experiences can be delivered in different ways and can perhaps be structured in a manner that is more effective. Examples of experiences that may benefit from insights about summary evaluations include entertainment events, service encounters (cf. Carmon et al. 1995), medical procedures (cf. Ariely 1998, Redelmeier and Kahneman 1996), and possibly even how people prefer their lives to end (cf. Diener, Wirtz and Oishi 2001). There may even be ways to write papers that leave their readers with very favorable overall assessments.

References

Ariely, Dan. 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.

----------, and Ziv Carmon. 2000. "Gestalt Characteristics of Experiences: The Defining Features of Summarized Events." *Journal of Behavioral Decision Making*, 13(2): 191-201.

----------, Daniel Kahneman, and George Loewenstein. 2000. "Joint Commentary on The Importance of Duration in Ratings of, and Choices Between, Sequences of Outcomes." *Journal of Experimental Psychology: General*, 129(4): 524-529.

----------, and George Loewenstein. 2000. "When does Duration Matter in Judgment and Decision Making?" *Journal of Experimental Psychology: General*, 129(4): 508-523.

----------, and Gal Zauberman. 2000. "On the Making of an Experience: The Effects of Breaking and Combining Experiences on their Overall Evaluation." *Journal of Behavioral Decision Making*, 13(2): 219-232.

----------2002. "Differential partitioning of Extended Experiences" Working paper Massachusetts Institute of Technology, 77 Mass Ave Cambridge MA 02139.

----------. 2001. "Seeing Sets: Representation by statistical properties." *Psychological Science* 12(2): 157-162.

Bodner, Ronit, and Drazen Prelec 2001. "Self-Signaling and Diagnostic Utility in Everyday Decision Making." In *Collected Essays in Psychology and Economics*, edited by Isabelle. Brocas and Juan. Carillo. Oxford University Press, 2001,

Carmon, Ziv, J. George Shanthikumar, and Tali F. Carmon .1995. "A Psychological Perspective on Service Segmentation: The Significance of Accounting for Consumers' Perceptions of Waiting and Service." *Management Science,* 41(11): 1806-1815.

----------, and Daniel Kahneman. 2000. "The Experienced Utility of Queuing: Real Time Affect and Retrospective Evaluations of Simulated Queues." Working paper, INSEAD, Marketing Department, Bd. de Constance, 77305 Fontainebleau Cedex, France.

Chapman, Gretchen B. 2000. "Preferences for Improving and Declining Sequences of Health Outcomes." *Journal of Behavioral Decision Making*, 13(2): 203-218.

----------. 1996. "Expectations and Preferences for Sequences of Health and Money." *Organizational Behavior and Human Decision Processes*, 67: 59-75.

Dhar, Ravi, and Klaus Wertenbroch. 2002. The Costs and Benefits of Temptation: Choice Set Effects on Consumption Utility. Working paper, INSEAD, Marketing Department, Bd. de Constance, 77305 Fontainebleau Cedex, France.

Diener, Ed, Derrick Wirtz, and Shigehiro Oishi. 2001. "End Effects of Rated Life Quality: The James Dean Effect." *Psychological Science*, 12(2): 124-128.

Ferster, C. B. & Skinner, B.F. 1957. "Schedules of Reinforcement." New York, NY: Appleton-Century-Crofts, Inc.

Fredrickson, Barbara L. 1991. "Anticipated endings: An Explanation for Selective Social Interaction." PhD. dissertation, Stanford University.

----------, and Daniel Kahneman. 1993. "Duration Neglect in Retrospective Evaluations of Affective Episodes." *Journal of Personality and Social Psychology*, 65: 45-55.

----------. 2000. "Extracting Meaning from Past Affective Experiences: The Importance of Peaks, Ends, and Specific Emotions." *Cognition and Emotion*, 14(4): 577-606.

Huber, Joel, John G. Lynch Jr., Kim Corfman, Jack Feldman, Morris Holbrook, Don Lehmann, Bertrand Munier, David Schkade, and Itamar Simonson. 1997. "Thinking About Values in Prospect and in Retrospect: Maximizing Experienced Utility." *Marketing Letters*, 8(2): 323-334.

Hsee, Christopher K., and Robert P. Abelson. 1991. "Velocity relation: Satisfaction as a Function of the First Derivative of Outcome over Time." *Journal of Personality and Social Psychology*, 60: 341-347.

----------, Peter Salovey, and Robert P. Abelson. 1994. "The Quasi-Acceleration Relation: Satisfaction as a Function of the Change of Velocity of Outcome Over Time." *Journal of Experimental Social Psychology*, 30: 96-111.

----------, George Loewenstein, Sally Blount, and Max Bazerman. 1999. "Preference Reversals Between Joint and Separate Evaluations of Options: A Theoretical Analysis." *Psychological Bulletin*, 125 (5), 576-590.

Kahneman, Daniel, and Dale T. Miller. 1986. "Norm Theory: Comparing Reality to its Alternatives." Psychological Review, 93: 136-153.

----------. 1992. "Reference Points, Anchors, Norms, and Mixed Feelings." *Organizational Behavior and Human Decision Processes*, 51: 296-312.

----------, Barbara L. Fredrickson, Charles A. Schreiber, and Donald A. Redelmeier. 1993. "When more Pain is Preferred to Less: Adding a Better End." *Psychological Science*, 4: 401-405.

----------, Peter P. Wakker, and Rakesh Sarin. 1997. "Back to Bentham? Explorations of Experienced Utility." *Quarterly Journal of Economics*, 112: 375-405.

----------. 2000. "Evaluation by Moments: Past and Future." In *Choices Values and Frames*, edited by Daniel Kahneman, and Amos Tversky. Cambridge Press: NY, NY.

Katz, J., Donald A. Redelmeier, and Daniel Kahneman. 1997, November. Memories of Painful Medical Procedures. Paper presented to the meeting of the American Pain Society, Washington, DC. and public health. San Diego, CA: Academic Press.

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

Langer Thomas, Rakesh Sarin, and Martin Weber. 2000. "The Retrospective Evaluation of Payment Sequences: Duration Neglect, Peak and End Effects." Working paper, Manheim University.

Loewenstein, George. 1987. "Anticipation and the Valuation of Delayed Consumption." *Economic Journal*, 97: 666-684.

----------, and Drazen Prelec. 1991. "Negative Time Preference." *American Economic Review: Papers and Proceedings*, 82(2): 347-352.

----------, and Elster J. (1992). Choice over Time. New York: Russell Sage Press.

----------, and Drazen Prelec. 1993. "Preferences for Sequences of Outcomes." *Psychological Review*, 100: 91-108.

----------. 1999. "Because it's There: The Challenge of Mountaineering… for Utility Theory." *Kyklos*, 52: 315-344.

Nowlis, Stephen M. and Itamar Simonson. 1997. "Attribute-Task Compatibility as a Determinant of Consumer Preference Reversals." *Journal of Marketing Research*, 34: 205-218.

Read, Daniel and Melanie Powell. Forthcoming. "Reasons for Sequence Preferences." *Journal of Behavioral Decision Making*.

Redelmeier, Donald A., and Daniel Kahneman. 1996. "Patients Memories of Painful Medical Treatments - Real-Time and Retrospective Evaluations of Two Minimally Invasive Procedures." *Pain*, 66: 3-8.

Ross, Willliam T. Jr., and Itamar Simonson. 1991. "Evaluations of Pairs of Experiences: A Preference for Happy Endings." *Journal of Behavioral Decision Making, 4*: 273-282.

Roth, Alvin E., Murnighan, J.K., and Schoumaker, Francoise. 1988. "The deadline effect in bargaining: Some experimental evidence. " *American Economic Review*, 78 (4) September, 806-23.

----------, and Ockenfels, Axel. Forthcoming. "Last-minute bidding and the rules for ending second-price auctions: Evidence from eBay and Amazon auctions on the internet." *American Economic Review*.

Schreiber, Charles A., and Daniel Kahneman. 2000. "Determinants of the Remembered Utility of Aversive Sounds." *Journal of Experimental Psychology: General*, 129: 27-42.

Sonnenschein, Bonnie.H., & Peter Shizgal. 2001 "Preference Reversals Displayed by Laboratory Rats Working for Rewarding Brain Stimulation." *Society for Judgment and Decision-Making*, Orlando, FL.

Trope Yaacov, and Nira Liberman. 2000. "Temporal Construal Theory of Time-Dependent Preferences." In *Economics and Psyc*hology, edited by Isabelle Brocas, and Juan D. Carrillo. Oxford University Press: NY.

Tykocinski Orit E. 2001. "I Never Had a Chance: Using Hindsight Tactics to Mitigate Disappointment." *Personality and Social Psychology Bulletin*, 27: 376-382.

Varey, Carol A., and Daniel Kahneman. 1992. "Experiences Extended across Time: Evaluation of Moments and Episodes." *Journal of Behavioral Decision Making*, 5: 169-185.

Author Notes:

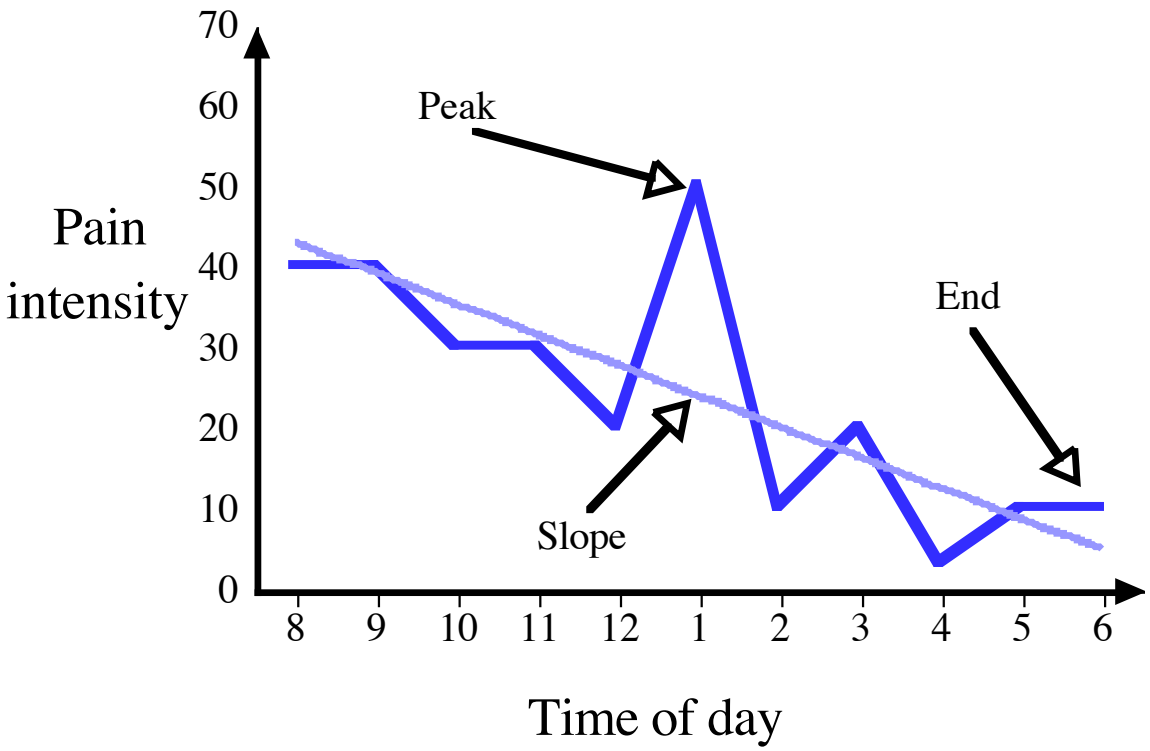
Dan Ariely is Sloan Career Development Professor at MIT, where he is affiliated with the Sloan School of Management and the Media Laboratory. He can be contacted at the following address: MIT, E56-329, Cambridge, MA 02142, or via e-mail at ariely@mit.edu. For more information about him see: web.mit.edu/ariely/www

Ziv Carmon is Associate Professor of Marketing at INSEAD. He can be contacted at the following address: INSEAD, Marketing Department, Bd. de Constance, 77305 Fontainebleau Cedex, France, or via e-mail: ziv.carmon@insead.edu. For more information about him see: www.insead.edu/facultyresearch/marketing/carmon/

Figure Captions

Figure 1: An example (based on Ariely and Carmon 2000) of an experience profile and three of its gestalt characteristics, based on the data of Subject 17 in the hospital study (assessment of a painful day at a hospital). *Peak* is the maximum intensity, *end* is the intensity at the final moment of the experience, and in this case *slope* is a single measure of the profile's overall linear trend (in gray).

Figure 2: Responses for the four experiments of Ariely and Loewenstein (2000), plotted separately for each experiment, and each duration. The four experiments were: 1) Ratings: ratings overall annoyance on a 0-100 scale. 2) Standard: ratings overall annoyance on a 0-100 scale, relative to a constant known standard that was 50. 3) WTA: minimum willingness to accept payments (¢) in exchange for the sounds. Choice: choice of each sound relative to a constant known standard. The measures are plotted in the original response scale. Mean annoyance on a 0-100 scale for the Ratings and Standard experiments. A monetary scale (¢) for the WTA experiment and the proportion of choice of the standard over the focal stimuli in the choice experiment.


70
60
50
40
30
20
10
0
Pain intensity
Peak
End
Slope
8
9
10
11
12
1
2
3
4
5
6
Time of day

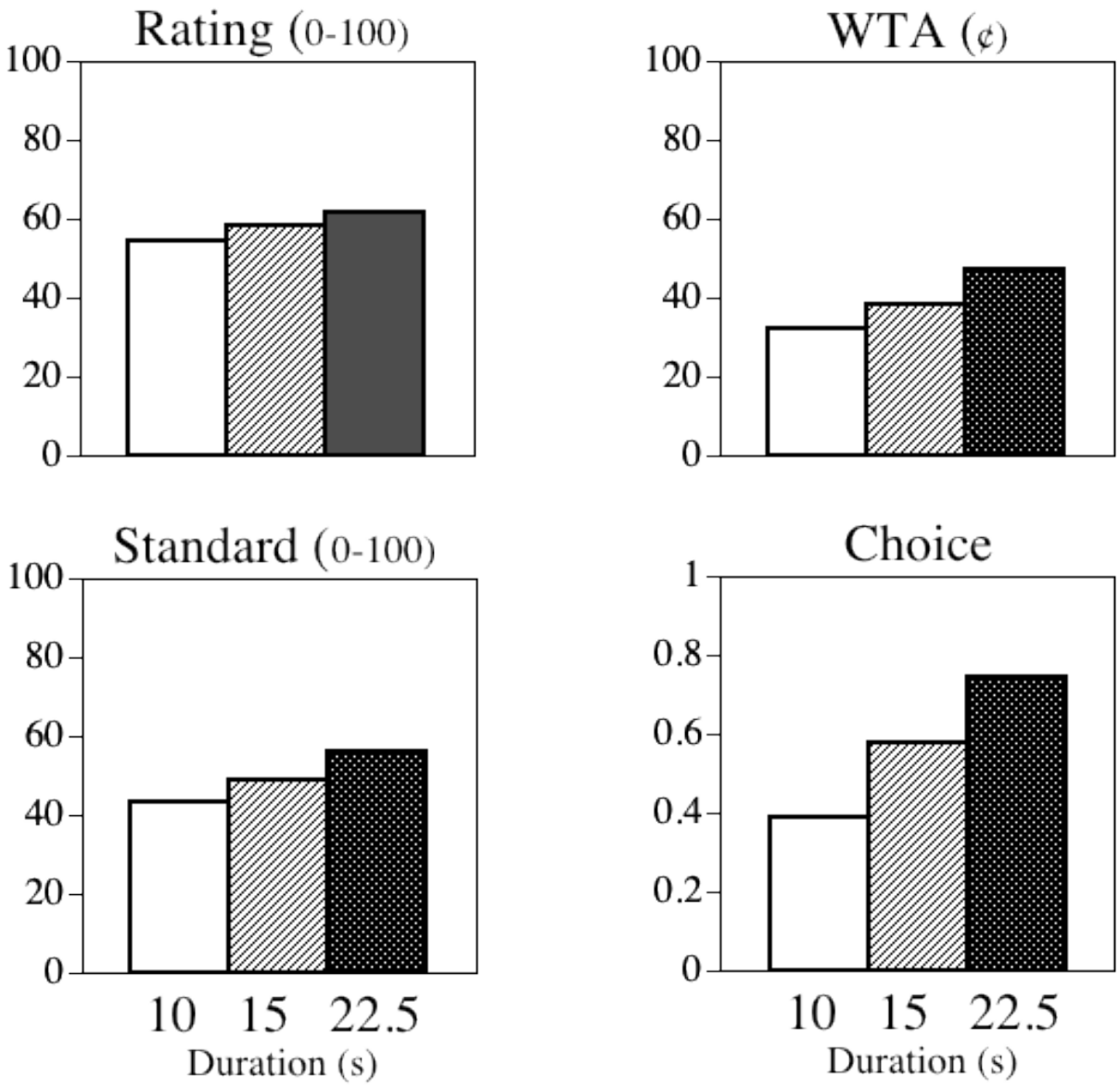

Rating (0-100)
100
80
60
40
20
0
WTA (¢)
100
80
60
40
20
0
Standard (0-100)
100
80
60
40
20
0
10 15 22.5
Duration (s)
Choice
1
0.8
0.6
0.4
0.2
0
10 15 22.5
Duration (s)

Table 1: Summary of four experiments in A&L (2000)

	Rating	Decision
Separate evaluation	Experiment #1 (Separate Ratings)	Experiment #2 (WTA)
Comparative evaluation	Experiment #3 (Rating relative to standard)	Experiment #4 (Choice)

[1] The ideas described in this paper draw on work we have conducted in this domain over the past few years. We draw heavily on ideas developed in Ariely and Carmon (2000) and Ariely and Loewenstein (2000).

[2] Ariely (1998) notes that it is interesting to consider whether people have both momentary and global evaluations and how independent they are.