THE EXPERIENCED UTILITY OF QUEUING: REAL-TIME AFFECT AND RETROSPECTIVE EVALUATIONS OF SIMULATED QUEUES

Ziv Carmon, Duke University

Daniel Kahneman, Princeton University

1995

This working paper is based on an essay from Carmon's PhD dissertation, which was conducted under the supervision of Kahneman, while both were at the University of California at Berkeley.

THE EXPERIENCED UTILITY OF QUEUING: REAL-TIME AFFECT AND RETROSPECTIVE EVALUATIONS OF SIMULATED QUEUES

Experiences such as service encounters or acts of consumption can be assessed either in real-time or retrospectively, reflecting the actual experienced utility and the way it summarized after-the-fact, respectively. To illustrate the potential for consumer research of a framework that combines real-time moment-to-moment recordings of affect with retrospective global evaluations, we explore these two types of evaluations and the relation between them in the context of simulated experiences of waiting-in-line. Specifically, we study the effect of two basic characteristics of a waiting line, its remaining length and its speed, on responses to queuing episodes. Subjects interacted with a computer program that graphically represented a queue in which the subject's position was indicated; the program also elicited affective responses in real time and global evaluations retrospectively. Four factors explained real-time affective responses: a strong response to the initial length of the queue, a positive change with each movement, a steady deterioration of affect between movements, and a trend of improving affect toward the end in queues that move rapidly. Unlike other types of events that have been studied in this paradigm, retrospective evaluations of the queuing experiences were accurately predicted by the affect recorded at their termination. We conclude with an extended discussion of experienced utility and consumer service queues, as our more general intention is to draw attention of consumer researchers to these two topics.

Consumer experiences, such as service encounters and product consumption, consist of events that extend over a period of time. Each extended event is associated with a stream of affective states that may vary in intensity and even in sign from moment to moment within a single episode. A summary evaluation is attached to events in past experiences: we remember them as more or less good or bad, and are accordingly prone to seek or avoid repeated encounters, or to recommend or not recommend the experience to others.

Momentary affect and global retrospective evaluation are distinct aspects of the experienced utility of an event (Kahneman, 1994). They can be operationalized respectively as (i) a moment-by-moment real-time record of affect during the experience; (ii) a global evaluation of the event. Each measure is of interest on its own for analysis of consumer behavior. Momentary affective responses represent the actual experience, and are likely to control immediate approach and avoidance decisions. In a service setting, for instance, negative momentary affect may result in reneging (leaving a queue), or balking (not joining a queue). Retrospective evaluations, on the other hand, summarize what is learned from the experience, and may affect future actions. Examples include word-of-mouth, complaint behavior, and the likelihood of future visits to an establishment. The relationship between these measures raises a question of substantial interest: how do consumers combine memories of a sequence of pleasant and unpleasant moments to form an evaluation of the experience as a whole?

In this paper we illustrate the potential for consumer research of a framework that combines moment-to-moment recordings of affect with retrospective global evaluations. This framework can serve to study both the dynamics of momentary affect and the rules that govern overall retrospective evaluations of episodes (see also Fredrickson and Kahneman, 1993; Kahneman et al., 1993). We demonstrate the value of this framework by using it to investigate customers' reactions to the experience of waiting in line.

Waiting-in-line can elicit powerful emotions that may vary significantly during the episode. When a queue appears to stall, for example, customers can become very frustrated, whereas unexpectedly rapid movement of the queue can actually be pleasurable. Hence, we believed that

the affective profiles of queuing may be sufficiently complex to warrant an investigation with this framework. Moreover, queuing is a frequently experienced by consumers who must routinely wait before receiving products and services they wish to acquire (in fact, it was estimated that on average, Americans spend 30 minutes a day waiting; Lindley 1988). The waiting portion of a service encounter can significantly affect customers' overall satisfaction with the transaction (see e.g., Folkes et al. 1987, Taylor 1994). Surprisingly, there has been little empirical research on queues (see Carmon, 1991; and Taylor, 1994, for a review of exceptions), despite growing recognition of the need for such research (Carmon et al., 1995; Larson, 1987).

We report two of the studies that we conducted with this general methodology. In Study 1, we use the technique to explore the effect of two basic characteristics of the queue, its remaining length and the speed at which is progresses, on momentary affect. We also investigate the relationship between momentary affect and global retrospective evaluations. Study 2 examines the effects of violated expectations on real-time affect and on global evaluations. Both studies were repeated with important variations of method, in order to examine the robustness of the technique's results. We conclude the paper with a discussion of the implications of our research, and opportunities for future research.

The Simulated Queuing Environment that we Studied

Focusing our investigation on the remaining length of queues and their speed required controlling for other situational factors. Examples of such characteristics that may influence the experience of queuing include physical dis/comfort, and social factors such as interpersonal comparisons, perceived fairness, and interactions with other customers.

We therefore developed an experimental environment in which queuing was graphically simulated on the screen of a personal computer (for successful use of simulation in consumer research of similar nature, see Bateson and Hui, 1992). A sample screen is illustrated in Figure 1. Subjects are asked to imagine that the long purple slanted line represents a waiting area, and that each short vertical black line represents a waiting person. The queue of short black vertical lines

is oriented to a service area (depicted by the short horizontal blue line) in which one person can be served at a time. When a 'customer' leaves the service area the entire queue moves forward. Subjects are instructed to identify the thick green line, which joins the queue shortly after the episode begins, as representing their own position in the queue. Their goal is to reach the head of the line, enter the service area, be served and leave, terminating the queuing episode.

In the next section we consider ideas about the factors that may determine momentary affective responses to and retrospective evaluations of the queues such as the ones we simulated.

Determinants of the Experienced Utility of Queuing

Experiences of waiting come in two main varieties, predictable waiting and open ended waiting. Predictable waiting, as in a Japanese train station, allows one to precisely predict the remaining duration of the wait. Open-ended waiting, as in waiting for a relative's surgery to end, is characterized by little or no feedback about the progress of the wait or its remaining duration.

Queuing, as in waiting for service at an airline counter, normally shares elements of both predictable and open-ended waiting. The remaining length of the queue and its observed rate of progress provide clues about time remaining but these clues are typically infrequent, unreliable and difficult to asses. We therefore believed that momentary affective responses to queuing would have elements of both types of waiting, as discussed below.

Momentary Affect

Predictable waiting: In predictable waiting the remaining duration can be perfectly anticipated at any given moment. We therefore generally expect aversiveness to peak initially and then diminish steadily as the goal is approached. Thus, a wait may not be aversive when you know it is about to end, but the further you are from the goal the more miserable you may feel. The reasoning is that affect is linked to expectations about the time remaining until the goal is reached. This idea entails a general improving trend of affect as the goal of release is approached. The improving trend can be accelerated, slowed, or even reversed, if the queue moves faster or

slower than expected. Note that we are attributing current affect to the anticipated duration of the remaining wait at that time, rather than to events in the past or in the present. The idea that expectations, like consumption, actually influence affective experiences is in the spirit of recent analyses of anticipatory emotions such as savoring and dread (Loewenstein, 1987; Loewenstein and Elster, 1992).

Open-ended waiting: In open-ended waiting, we expect steady escalation of negative affect. Hence when customers have no clear information about the remaining waiting duration they become increasingly frustrated over time. This idea is consistent with Osuna's (1985) analytical model of waiting which suggests that the aversiveness of waiting escalates over time, as a result of what he names the *psychological cost of waiting*.

Dual Response to Waiting: Since waiting in line has elements of both types of waiting, we propose a dual-response idea to predict the momentary affective responses it invokes. This hypothesis accepts the idea that consumers become gradually demoralized as they wait, but also incorporates a positive affective response to each movement of the queue (see also the *Renewable Resource Model*, Linville and Fischer, 1991).

We expect the initial responses to queuing to be determined by expectations of the remaining duration, consistent with the predictable waiting component in queues. More specifically, the initial observation of the length of the queue that one will join will determine the initial affective response. The logic is that a queue's length can be immediately assessed whereas its speed cannot and subjects will therefore use the length of the queue as a proxy for the expected duration. Note that the open-ended component of queuing does not affect the initial responses as it develops over time.

The dual-response idea further predicts subsequent cycles in which a positive affective response to each forward movement of the queue is followed by a negative trend, which is only interrupted by the next advance. The general trend of the affective profile over the queuing

episode is difficult to predict as it will be determined by the balance of the two opposing responses. Note that the idea entails progressive escalation of negative affect if a queue inexplicably fails to advance, and predicts consistent improvement (after the initial response) in the extreme case of a queue that provides continuous and clear evidence of progress. Responses to these two extreme cases are thus consistent with the responses to "pure" forms of open-ended and predictable waiting.

Global Retrospective Evaluations

Analyses of waiting in Operations Research commonly adopt an economic framework (see, for example, the *Economic Cost Model*, Nichols et al., 1977), in which the disutility of waiting is measured by the opportunity cost of the agent's time. The economic value of time varies with the income of different individuals, and it may vary across situations for any individual (e.g., missing an hour of work or an hour of leisure). However, the value of time is unlikely to vary in the course of a specific waiting incident. An economic analysis therefore predicts that, for a given person, the global evaluation of episodes of waiting should depend only on their total duration. This analysis appears inadequate, normatively as well as descriptively, because it ignores the social and psychological factors that determine the meaning of waiting.

If time alone does not govern global evaluations of waiting episodes, what does? Our theoretical framework suggests that retrospective evaluations should be analyzed in terms of characteristics of the real-time affective responses. Recent studies of experienced utility indicate that retrospective evaluations are predicted with substantial accuracy by a weighted average of the real-time affect reported at the worst moment of the experience (the peak, hereafter), and at its end. This has been shown for ratings of fictitious experiences (Varey and Kahneman, 1992) as well as for actual experiences of pleasant and aversive films (Fredrickson and Kahneman, 1993), a cold-pressor situation (Kahneman et al., 1993) and medical procedures (Redelmeier and Kahneman, 1995). It has also been shown in the context of affective responses to advertisements (Baumgartner et al., 1995).

Queuing differs from these settings in that it is typically an instrumental, goal-directed activity. Furthermore, the occasions that evoke affect are unusual: the negative affect of queuing is associated with non-events such as expectations (the initial response to long queues), or with their discomfirmation (when the queue moves more quickly or more slowly than anticipated). As a result of these differences we expect global retrospective evaluations to be predicted by the affect at the end of the experience. Recall that this would be unlike the types of experiences that have been previously studied, in which these evaluations were predicted by a combination of affect at the end and at the peak.

The logic is that in goal-oriented activities, people may retrospectively discount the significance of the process, and be influenced by the affect experienced when the goal is reached (or perhaps abandoned). Moreover, if the main sources of affect during an episode are primarily influenced expectations and disconfirmation (whether positive or negative) the emotions associated with such non-events may seem irrelevant, when the uncertainty is resolved at the end of the episode. In summary, we expect that the global evaluations of queuing experiences will be predicted by end-affect.

General Method

Our program exposed subjects to several separate simulated queuing experiences during an experimental session of 60-90 minutes. The queuing episodes differed in the number of 'people' (represented by short black lines in Figure 1) that were in line ahead of the subject, and in the speed at which the queue progressed. This was possible since the queues in our experiment progressed at a much faster pace than most queues in the real world. The average duration of the stay of an individual at the service counter was 16 seconds in some experiments, 24 in others, and the total duration of queues in our experiments ranged from 80 sec to 288 sec.

An "affect meter," which is shown at the bottom of the screen displaying the program (see Figure 1) allowed us to measure momentary affect. Subjects used the two arrow keys on the keyboard to control the direction and length of a row of arrows displayed below the queue (see

Figure 1). The intensity of negative or positive feelings was indicated respectively by the number of arrows (out of a maximum of 10) pointing left or right. Neutral affect was indicated when the vertical line in the middle of the scale appeared alone. For example, the record shown in Figure 1 would be scored -6. Similar scales have been shown to be valid in studies of the response to advertisements (see e.g., Hughes, 1992) and in studies of emotionally loaded interpersonal interactions (e.g., Gottman et al., 1985). Pretests indicated that subjects found our simulated queues compelling, and readily fell in with the instruction to monitor and report their feelings of minor irritation or pleasure.

A queue we named Normal, which progressed at a constant pace, was included in each experiment to provide an implicit standard of comparison for other episodes. The Normal queue was used for practice before beginning actual data collection, was repeated as the first episode in the experiment proper, and was repeated again half-way through the session and also at its end. To control for the possibility that evaluations of a queue other than Normal would contaminate that of another queue, we constructed two different sequences of the queues for each experiment. Two additional sequences were produced by inverting the original orders, yielding a total of four different sequences of queues. Subjects were randomly assigned to one of the four sequences.

To reduce direct effects of the real-time ratings on retrospective evaluations, subjects performed an arithmetic task for 25 seconds at the end of each queuing episode. At the end of this interval, they rated the overall experience of the preceding episode on a 21-point scale ranging from 'very negative' to 'very positive.' Subjects were then asked a trivia question which occupied roughly 2-3 minutes (e.g., "list all the European capitals you can think of")[1]. After completing this task subjects proceeded to the next episode in their sequence of queues.

[1] This task was originally intended to eliminate carryover effects across queuing episodes. However, informal observations suggested that subjects liked the trivia question, and interpreted it as a goal that would be reached at the end of the waiting episode. Fortuitously, this filler task may have served to make the waiting situation more real for subjects.

Study 1

Recall that in Study 1 we explore how the remaining length of the queue and the speed at which it progresses, influence the time-course of affect during queuing, and also study the relationship between real-time affect and global retrospective evaluations. Also recall that we expected: (1) The initial real-time response to be a function of the initial queue length. (2) A saw-tooth-like affective profile, consisting of a decline in affect until the queue advances, interrupted by a brief positive response to the movement of the queue. (3) The global evaluations to be predicted by affect at the end of the episode.

We examine the responses to six different types of queues. A queue named Normal was the implicit standard against which we expected our subjects to compare the other queues. Three episodes, Fast, Slow and Empty, varied in initial length and in speed, with total duration constant. Two episodes, Long and Short, differed in the initial length of the queue and in the duration of the episode, with movement speed constant.

There were two different versions of the study, which we label Experiments 1A and 1B. In Experiment 1A, the modal interval between movements of the queue (the normal service time) was 16 seconds. This was also the constant speed of the Normal queue for that experiment. Experiment 1B was identical in all respects to Experiment 1A, but its basic rhythm was slower: all time intervals in Experiment 1A were multiplied by 1.5 to generate the slow version of the experiment. The aim of this manipulation was to find out whether the build-up of impatience and other affect in our experimental situation was dependent on the basic rhythm of the experiment. Finding major differences between Experiment 1A and 1B would raise severe doubts about the generalizability of results to real life queues, which typically move much more slowly than the queues of our simulations.

Method

***Subjects*.** Subjects were paid $5 per hour for a session lasting 60-90 minutes. Their ages ranged from 19 to 53. There were 18 subjects in experiment 1A, and 25 in experiment 1B.

***Procedure*.** The length of the Normal queue was 10 'people' and its speed was one advance per 16 seconds in Experiment 1A, and one advance per 24 seconds in Experiment 1B. This ratio was maintained across the two experiments. The Short and Long queues shared the speed of the Normal queue, but they differed in the initial length of the queue (5 or 15 ‘people’, respectively) and in its duration. The Fast and Slow queues differed from the Normal queue both in initial length (20 or 5 vs. 10 ‘people’ in the Normal queue) and in speed (8 or 32 seconds between movements, vs. 16). In the Empty queue, there was a person at the service area, with no other 'customers' ahead in the queue, and the subject was kept waiting for a relatively long time (144 sec in Experiment 1A) before entering the service area.

Results & Discussion -- Study 1

Before discussing the substance of this study's results, we conduct a preliminary analysis to examine aspects of their reliability and robustness.

Preliminary Analyses. The average results that are depicted in Charts 1 and 2 are quite representative of the data of individual subjects. To assess the concordance of subjects we correlated each subject's responses on successive seconds of each episode with the mean of the responses of all other subjects. The mean of these correlations was 0.90 in Experiment 1A and 0.82 in Experiment 1B. The consistency of affective responses permits us to ignore individual differences and to focus our analysis on group averages.

Order effects in the response to the queues were examined in two ways. First, we compared subjects' responses to the three repetitions of the Normal queue (see Chart 1). We also used between-subject analyses to compare the responses to other sessions, which some subjects encountered early in the experiment and others encountered late. None of these analyses revealed significant differences that could be attributed to sequence effects. We conclude that subjects adapted quickly to the experimental situation. Valid expectations and stable criteria for affective

responses were apparently established by the Normal queue, which subjects encountered both in the pre-experimental practice and in the first episode in the session. Hence order effects did not seem to be a problem in this experiment.

To summarize, the method yielded generally encouraging results in this study: they demonstrated high concordance among subjects, and appropriate sensitivity of affect reports to differences among the simulated queues.

***Analysis of Real-Time Responses*.** We now turn to a more detailed inspection of the pattern of affective changes that occurred in the various queues. We will focus on the results of Experiment 1A. The results of Experiment 1B were very similar, though somewhat noisier.

The early responses to the queues are negative, except for queues that are markedly shorter than Normal (see Charts 1 and 2). The initial affective response appears to be determined by the initial length of the queue, as predicted. The correlation between the log of queue length and early affect (see Table 1) is 0.94 for both Experiments 1A and 1B. The initial affective response to the length of the queue can only be an expectation effect, because it occurs before there is time for significant costs or pleasures to accumulate. By their use of the affect scale, the subjects indicate that the expectation of some degree of annoyance (or sometimes pleasure) is not only a cognitive event; expectations of affect are affective events in their own right. Subjects appear to be "consuming" their expectations of future pleasure or displeasure. Displeasure is associated with the expectation of waiting, but the expectation of waiting much less than usual is not only less unpleasant -- it is an occasion for mild pleasure. The initial responses we found are thus consistent with the dual-response idea.

Subsequent responses to queuing during the middle phase also appear consistent with the dual-response idea. As illustrated most clearly in Chart 2, the response to continued waiting is a steady decline of affect, as proposed by Osuna (1985). This decline is reversed by a positive change following each advance of the queue. The resulting saw-tooth pattern is the prominent feature of the response to the Slow queue (Chart 2). It is tempting to attribute the pattern of Chart

2 to the disappointment that subjects experience when they detect that the queue has failed to move at the normal time. However, the data of Chart 1 rule out this idea: scallops corresponding to individual movements of the queue are found in the response to the Normal queue, which moves precisely as expected. As proposed in the dual-response idea, there is a steady deterioration of affect while waiting, which is temporarily reversed by the pleasure that is associated with each movement of the queue.

During the last phase of the queue, the positive response to each movement appears to become much larger than the negative response that occurs between movements: The pleasure of moving forward is evidently enhanced when the goal of release is seen to be near. The altered balance between positive and negative responses creates a steep improvement of affect, which is perhaps the most salient feature of Charts 1 and 2. With the exception of the Empty episode, all queues end on a definitely pleasurable note. Thus, our subjects described the approach to the goal as an experience of increasing pleasantness. While the steep improvement in the last stage is not a feature we specifically expected, it is consistent with the dual response idea. The changes of affect during the last 32 seconds are strikingly similar for the five queues that move at Normal speed (see Chart 1), consistent with the dual-response idea. Another observation, which we examine in Study 2, is that the size of the affective response in this terminal phase may be determined by the speed of motion.

The results of Chart 1 are relevant to a test of the role of expectations component in real-time affect. The three queues for which results are shown differed in length and duration, but not in speed. If affect throughout the queue were determined exclusively by expectations about the time remaining, it would be possible to superimpose the final phases of the three records. This is clearly not the case: the affective differences between long, normal and short queues are maintained to the end. The expectations component of the dual-response explains the initial differences between the responses to these queues, and the combination of the two types of responses explains subsequent changes of affect. The outcome is a substantial effect of initial

queue length on the affect recorded at the end of the queuing episode. Subjects indicate more pleasure at the end of short queues than at the end of long ones ($F(2,17) = 31.6$, $p<.01$).

Recall that the two experiments of Study 1 differed only in basic rhythm: every temporal interval in Experiment 1A was increased by half for Experiment 1B. Our objective was to find out whether subjects responded to the absolute lapse of time, or whether they compared the rhythm of events in a particular queue to a norm established earlier in the session. For each type of queue we matched the mean response observed on each second of Experiment 1A to the mean response observed at the corresponding time in Experiment 1B. For example, the response recorded 10 seconds after the onset of an episode in Experiment 1A was matched to the response recorded 15 seconds into the same episode in Experiment 1B. Chart 2 shows the response to two queues (Empty and Slow) in Experiment 1A, and the compressed response to the same conditions in Experiment 1B. We also computed correlations between the profiles for each queue in the two experiments. The highest correlation was for the Empty queue (0.98); the lowest was for the Slow queue (0.78). Over all of the episodes included in the experiment, the mean of these correlations was 0.94. This suggests that the affective responses observed in our experiment were mainly controlled by relative rather than by absolute time. For example, subjects in the Long queue in Experiment 1A probably expected to wait no longer than ones in the Normal queue of Experiment 1B (the actual duration of both was 240 seconds). However, the mean initial response to the former was more negative than the corresponding reaction to the latter (-2.94 vs. -0.76; $t(38) =$, $p< .01$). Affect was slightly more negative in Experiment 1B, where queues lasted uniformly longer. Overall, the difference between the average affect recorded in the two experiments was small (0.9 scale units; $t(41) = 3.8$, $p< .01$). This correspondence between experiments varying in basic rhythm supports the generalizability of our results to real queuing environments, which do not typically progress as quickly as our simulated queues.

To summarize, the findings of Experiment 1 are compatible with a relatively simple four component model of the affect observed at any time during queuing episodes (which is also consistent with the dual-response idea we proposed): (1) an initial response to the length of the

queue, which reflects the affect associated with the expectation of future affect; (2) a negative response to waiting; (3) a positive response to each advance of the queue (for the Normal queue this response is slightly larger than the negative change during the interval between movements); and (4) a substantially enhanced positive response to each advance when the queue is short. These four factors appear to be additive. For example, the influence of initial queue length is not dissipated in the course of queuing, as illustrated by Chart 1.

Chart 3 presents the results of two conditions of a separate study, conducted in the same general method, which provides further evidence for the dual-response idea of affect during queuing. Data were collected from 15 subjects. The Normal queue consisted of 12 'customers' who waited 16 seconds between movements. The queue labeled Dom consisted of 18 customers. The first 12 were served every 16 seconds, the last 6 were served every 3 seconds. The components of the affective response we observed in the other queues are clearly evident: (1) The initial response to the longer queue (Dom) is more negative than the response to the Normal queue. (2) A negative response is evident in the intervals between advances of the queue. (3) The positive response to advances is at first similar for the two queues; it increases progressively when the queues become shorter, causing an initial divergence between the curves for the two conditions. (4) The enhanced positive response is the dominant feature of the final phase of queuing. Because there is no time for any buildup of negative affect between advances when the queue moves quickly, the mean recorded affect is more positive at the end of the Dom queue than at the end of the Normal queue (7.1, and 4.9, respectively; ($t(14)= 4.7$; $p<.01$).

Global Retrospective Evaluations. Recall that the economic analysis predicts that, for a given person, the global evaluation of episodes of waiting should depend only on their total duration. It is immediately apparent that the results violate this prediction. For example, the Empty and Fast sessions, though of equal duration, were evaluated -2.3 and 1.2 on average ($t(14) = 3.7$, $p<.01$).

As we anticipated, the relationship between real-time and retrospective evaluations in this study is unlike the findings of several other studies within the same general paradigm. Recall that

in studies involving pleasant and unpleasant films (Fredrickson and Kahneman, 1993), immersion of the hand in painfully cold water (Kahneman et al., 1993) and the pain of medical procedures (Redelmeier and Kahneman, 1995), retrospective evaluations were predicted by a weighted average of the most intense affect recorded during the episode and of the affect recorded in its final moments. In the present study, however, only end affect seemed to matter. We based this prediction on the two ways in which queuing is different from these other experiences (affect during queuing is influenced primarily by non-events, and queuing is typically a goal-directed activity).

Specifically, the retrospective evaluations in this study are predicted quite accurately by the real-time affect recorded at the end of the different episodes. A rank-order correlation between these variables was computed within the data of each subject. The average correlation for 18 subjects was 0.75. The correlation between the averages of the final real-time responses and the average retrospective rating was of course higher ($r = 0.90$). The affect experienced while queuing apparently played little or no role in the retrospective evaluation of the experience, which was dominated by end-affect.

These conclusions also hold up in the conditions of the supplementary study shown in Chart 3. Although the Dom queue was longer than the Normal queue and took a longer time, both its end affect and its global evaluation were more favorable (the means of retrospective evaluations were 1.8 for Dom and 0.7 for the Normal queue, $t(14)=-2.1$; $p<.03$). Interestingly, while it is unlikely that consumers would voluntarily choose to wait longer, apparently in some situations they would end up feeling better if they did.

Study 2

An important methodological concern that we wanted to address in this study was a possibility that the measurement of real-time affect influenced the global retrospective judgments our subjects indicated. In other words, we intended to preclude possible contamination of the global evaluations by previous real-time responses. To this aim, the experiment was conducted in

two versions. Experiments 2A and 2B were identical in all respects, but the subjects in Experiment 2B did not report affect in real time.

Another objective was to confirm the relationship between the affect recorded at the end of the queues and their global retrospective evaluation. In particular, we wanted to examine whether the strong association between end-affect and global evaluations would persist if the end-affect was very different from the affect sensed during the bulk of the experience. For that purpose we compared responses to five queuing experiences, all matched for total duration, but differing in initial queue length and in speed of motion. Two queues were initially significantly longer than the Normal queue, but moved correspondingly faster. Two queues were both significantly shorter and slower than the Normal queue.

Method

Subjects. Subjects in this experiment were recruited in the same way as in Study 1. They were paid $5 for a session lasting roughly one hour. Fourteen subjects participated in Experiment 2A, and 13 in Experiment 2B.

Procedure. As in Experiment 1, subjects were randomly assigned to one of the four different sequences of queues. All subjects first practiced with the Normal queue, which was repeated on two other occasions. The Normal queue consisted of twelve 'customers' who were served within 24 seconds. The speed and number of people was twice and four times larger in the Long and Very Long queues, and twice and four times smaller in the Short and Very Short queues, respectively. Thus, the total duration of all five queuing episodes in this experiment was the same.

Results & Discussion -- Study 2

It is immediately apparent that the initial negative response to the queues was determined by their length, and that the final response was determined by their speed -- which in this study was inversely related to their length. This pattern is highly replicable. We computed for each subject a

correlation between (i) the log of the initial queue length and the preliminary affective response; (ii) the speed of the queue and affect recorded at the end of the episode. These correlations averaged -0.96 and 0.73, respectively. With total queuing duration constant, a longer queue that moves faster is associated with much better end affect than a shorter queue that moves more slowly.

The graphs of Figure 4 can be understood in terms of the rules that govern the sequence of affective changes from the initial sight of the queue to its termination. The sight of the short queues evoked a response of pleasure, graded by queue size (the peaks were 5.0 for the Very Short queue and 1.6 for the Short queue; $t(13)= 4.6$, $p<.01$). The initial response to the Very Long and Long queues is negative, and more negative for the Very Long than for the Long queue (-8.4 and -5.7, respectively, ($t(13)= 4.0$, $p<.01$). All five queues subsequently display the dual-response cycles, but those are particularly salient in the slower ones. In the longer queues the advances of the queues soon cause affect to improve because the movement is rapid and there is not always time for ratings to decline before the next movement initiates a new positive change. For the slow queues, the balance of positive and negative responses tips in the direction of the latter, especially for the Very Slow queues. The curves cross near the end of the episodes. End affect is more negative for the queue for which initial hopes had been higher, though the duration of both queues is the same.

Figure 4 also presents the global evaluations of the five queues. These global evaluations, like final affect, are correlated positively with queue speed -- and therefore correlated negatively with initial queue length. The correlation was high in the data of individual subjects (the mean rank-correlation of global evaluations with queue speed was 0.82 in Experiment 2A, and 0.89 in Experiment 2B). As we observed in Experiment 1, there is essentially perfect ordinal correspondence between the affect that subjects reported at the end of the queue and their subsequent evaluation of it. This result is especially striking in the present experiment, because the final ordering of the conditions by speed of movement is only evident in the affective response during the final seconds of these queues. The graphs in Figure 4 are dominated by initial expectations for much of their duration, and the average level of affect reported for the

entire episode is also ordered by queue length rather than by speed. Thus, the correlation between the global evaluations of queue and the average real-time affect is strongly negative in this experiment: the mean of within-subject rank-order correlations is -0.75). In contrast, the mean correlation between global evaluations and end affect was 0.86 in Experiment 2A.

It could perhaps be argued that the correlation between global evaluation and end-affect are due to contamination of the latter by the former. An important feature of Experiment 2 is that it allows us to reject this possibility, since the global evaluations were essentially the same in Experiment 2B, where no real-time ratings were made. We conclude that the instruction to report real-time affect did not substantially alter retrospective evaluations in this experiment. Similar conclusions were reached in several recent studies within the same general paradigm (Fredrickson and Kahneman, 1993; Kahneman et al., 1993). Thus, the present study offers three reassuring messages about the real time records: they appeared to be sensitive to the *level* of affect at any time and not only to affective changes, they predicted global evaluations with high accuracy, and they were not intrusive.

General Discussion

In this paper we develop and illustrate a framework and a technique for the study of experienced utility, which distinguishes retrospective evaluations from real-time affect, and explains the former in terms of the latter. To demonstrate the value of such a technique we used it to study some of the factors that determine responses to the experience of queuing, both in real time and in retrospection.

In the following sections we discuss implications and applications of some of our findings.

The Experience of Waiting in Line

Our subjects indicated by their responses that the expectation of an affectively relevant event can itself be a powerful source of affect. It may therefore be important to manage consumers' expectations, particularly when real-time affect is of particular concern. In our experiments, the

length of the queues, in particular, seemed to have a major effect on momentary affect. This may be a reason firms such as Disney, for example, disguise the length of waiting lines by having winding queues.

A result we had not specifically predicted was that our subjects began responding negatively to the failure of the queue to advance almost immediately after the pleasure of the preceding advance -- long before the next movement could be expected. Hence, it may be important to provide precise feedback about the progress of the queue even when customers know roughly how long they will have to wait. For example, an electronic map indicating the location of a scheduled train on its course may make waiting for it to arrive less unpleasant. The frequency at which such feedback should be provided is a question that should be studied further.

Service environments with multiple servers can either have a separate queue for each one (multiple queues), or combined queues for several servers that are longer and move faster (combined queue). Some organizations, such as banks, tend to have combined queues, whereas others, such as supermarkets, often have multiple queues. Other establishments, such as fast food chains, use a mixture of the two. Operations researchers have argued that combining queues is almost always preferable, because it results in the average expected waiting time not being longer, and often shorter (-- a major contribution of Operations Research to the 'real world'; Rothkopf et al. 1987). Researchers in this area have wondered why many queues are not set up in that way. We suggest that combining queues may be advisable if one is concerned with consumers' retrospective evaluations. However, multiple queues may be advantageous if one is concerned with how consumers feel while they are waiting, (see Rothkopf et al., for other reasons multiple queues may sometimes be beneficial).

More generally there is a very broad literature on queuing that relies on mathematical modeling (see e.g., Kleinrock 1975 for a review). Carmon, Shanthikumar, and Carmon (1995) argue that many of the standard queuing models must be modified to be applied directly to consumer service queues. Carmon et al. (1995) illustrate the importance of adjusting models to account for aspects of the psychological cost of waiting, showing that traditional prescriptions

can be inconsistent with those derived with adjusted models. Combining the framework presented in this paper with either a simulation technique like ours, or with other types of empirical work, could help significantly in this regard. The goal would be to assess the psychological cost in given queue settings, to provide queuing modelers information as to how adjustments should be made.

Finally, the effect of duration was not as dominant as is suggested by the Economic Cost Model, or by the traditional queuing literature. We suspect that duration might play a more important role for waits that are exceptionally long, although whether this is indeed true is a question that should be studied further. Our demonstration that the global evaluations can be improved when a fast moving segment of waiting is added at the end, thereby increasing the total duration is particularly damaging to this perspective. The high correlation between the affect recorded at the end of the queue and subsequent evaluations suggests that focusing on the final sensations may be the most effective way to improve the overall judgment of an event if resources are limited. More generally, concentrating on how queues are experienced may prove more cost effective than focusing on shortening the waiting duration (see Larson 1987 for nice anecdotes of this type).

Experienced Utility

Recent studies of experienced utility have shown that intuition is a poor guide in determining the aspects of the real-time experience that govern later evaluations (see for example Kahneman and Snell 1992). The most striking characteristic of the process that assigns global evaluations to experiences is its selectivity: several recent experiments have found support for a simple "Peak & End" rule, in which the retrospective evaluation of (aversive) experiences is accurately predicted a weighted average of the affective intensity at its worst and its last moment. The results of the present study supported an even simpler rule of evaluation: experiences of queuing appeared to be scored entirely on the basis of the sign and intensity of affect as the experience ended,

regardless of what went on before it. It is fair to describe remembered utility as a biased and incomplete representation of experienced utility.

The biases of remembered utility that we discussed here are not produced by forgetting. The selective process that assigns weight to some aspects of the experience and not to others appears to operate when the representation of an experience is first formed. Indeed, significant neglect of duration has been observed when subjects evaluated a description of another person's experience, a task that involved no episodic memory at all (Varey and Kahneman, 1992, Experiment 2). While distortions of hedonic memory may certainly occur over time, retrospective evaluations are suspect even when they are made immediately. Remembered utility should be understood in terms of rules of evaluation that turn an extended hedonic or affective experience into a single global score that assigns a value of attractiveness or aversiveness to the experience.

The rules of retrospective evaluations can yield unreasonable preferences. For example, most subjects preferred to repeat an episode of immersion of a hand in painfully cold water which ended with a gradual increase in the temperature of the water, rather than repeating a shorter episode which ended with pain at its peak (Kahneman et al. 1993). A similar result was reported here (see Chart 3): a long queue that ended with a very rapid advance elicited more positive affect and a more favorable retrospective evaluation than a shorter queue. These peculiarities of evaluation raise new problems and may create new opportunities. For example, it may be possible for a physician to endow the patient with a less aversive memory of a painful medical procedure by making it longer (but less painful during the additional period), thus increasing the total pain that the patient actually experiences (see Redelmeier and Kahneman, 1995).

More generally, the conception of experienced utility that emerges from this work is more complex than the standard notion of consumer satisfaction. The satisfaction literature often presents both expectations and disconfirmation as discrete constructs, with the former existing before the event and the latter after it. Our research suggests that both expectations and disconfirmation may be updated continuously as the event occurs (also see Wilton and Tse, 1988; Boulding et. al., 1993).

The dissociations between measures of real-time and remembered utility imply that no single index of satisfaction or utility will do. The objectives of optimizing experience and optimizing memory may lead to different choices. Such dilemmas are likely to be true for a variety of consumer services and product consumption experiences, and could raise subtle ethical issues. In many situations tastes (and therefore choices) will reflect remembered utility, rather than the actual experience, and marketers should therefore focus on improving the former. In this context, it is most natural to treat real-time experience as the independent variable and retrospective evaluations as the dependent variable which is to be predicted and explained. There are situations in which experienced utility may be more important. Examples include painful medical procedures, and settings in which consumers may balk or decide whether or not to wait depending on a momentary impression.

Where do the rules of retrospective evaluation come from? Which of these rules are biologically favored, which are culturally determined? Studies of the determinants of approach and avoidance tendencies in animals may help answer these questions. It is not unreasonable to speculate that "Peak & End" evaluation and duration neglect may be found to apply to the preferences of lower animals: in evaluating episodes of pain, the worst moment and the end of episode are likely to be singular in their adaptive significance, whereas the duration of the episode may matter little. Thus, the rules of evaluation may have the function of increasing inclusive fitness, even if they are not designed to minimize pain and maximize pleasure.

Other rules of evaluation may be culturally defined. The conception of the good life varies greatly across cultures, and this conception may affect the view of particular events. The rule of evaluation that was observed in the present study fits the familiar adage "All's well that ends well," which may be more representative of the beliefs in some western cultures. We do not know the range of situations to which this rule applies. Two plausible explanations, which are not mutually exclusive, were mentioned here. The first is that the "End rule" applies to situations of uncertainty, in which the main sources of affect during an episode are expectations, whether positive or negative. When the uncertainty is resolved at the end of the episode, the emotions

associated with expectations may seem irrelevant, and only the real outcome would count. Another hypothesis is that queuing is treated as an instance of a class of instrumental, goal-oriented activities, where the affect experienced when the goal is reached (or perhaps when the goal is abandoned) provides a joint evaluation of the costs and benefits of the activity and its outcome. Further research is required to test these hypotheses, which suggest that the affect associated with expectations and with instrumental activities is given little weight, except to the extent that it influences the affective response to the eventual outcome.

The prospects for research within this framework depend on the availability of a sensitive, simple and reliable measure of real-time affect, which is not overly intrusive in distorting the immediate experience, and does not contaminate subsequent global evaluations. In that sense, the picture of the affective response to queuing that emerged from the present work was encouraging. The high quality of the real-time record of affect was one of the salient results of our experiments. The responses were detailed, and remarkably coherent. The level of inter-subject agreement was high and the temporal resolution of the measure was adequate, as indicated by the rather precise responses to advances of the queues. Furthermore, the task of reporting real-time affect appeared to have no significant effects on retrospective evaluations: the determinants of these evaluations appear to be the same for subjects who reported their affect in real time and for subjects who did not (see also Fredrickson and Kahneman, 1993; Kahneman et al, 1993; Redelmeier and Kahneman, 1995).

An important question for the interpretation of the real-time records of affect is whether events are associated with levels of affect or with changes of affect. In other words, allowing for inevitable response lags, do subjects' ratings indicate the sign and intensity of affect they experience at each moment? The alternative hypothesis, which seems plausible, is that subjects use the keys to indicate the direction and intensity of *changes* of affect, without paying much attention to the *position* they indicate on the scale. The hypotheses are testable. If subjects are only sensitive to changes and ignore the position they mark on the scale, then their recorded responses should drift increasingly from any initial calibration because 'errors' in the response to

successive changes will not be corrected. If the responses are increasingly affected by accumulating noise, the variability of the responses recorded by different individuals should increase steadily in the course of the recording. In our experiments there was no general trend for these standard deviations to increase. This is not the result that would have been expected if subjects were merely pressing the arrow keys to indicate changes in their affect. We conclude that subjects also monitored the scale position they reported, to ensure that its correspondence with their momentary affective state. This conclusion implies, of course, that both affective level and affective change are measured with some validity in this technique.

The range of applications of the approach and the technique that have been illustrated here should readily extend to diverse areas of consumer psychology. Examples include cultural or media events such as concerts, films, commercials, books, and dining out. For such applications it will be useful to develop 'experience profiles' describing typical sequences of hedonic and affective experiences associated with different products and services, and to search for ways of improving these profiles and the retrospective evaluations that they generate.

References

Bateson, John E. G. and Michael K. Hui (1992), The Ecological Validity of Photographic Slides and Videotapes in Simulating the Service Setting, *Journal of Consumer Research*, 19, 271-281.

Baumgartner, Hans, Mita Sujan and Dan Padgett (1995), Patterns of Affective Reactions to Ads: The Integration of Moment-to-Moment Responses into Overall Judgments, Pennsylvania State University working paper.

Boulding, William, Ajay Kalra, Ricard Staelin and Valerie A. Zeithaml (1993), A dynamic Process Model of Service Quality: From Expectations to Behavioral Intentions, *Journal of Marketing Research,* 7-27

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, 1806-1815.

Carmon, Ziv (1991), "*Recent Studies of Time in Consumer Behavior*," in Hollman, Rebecca C. and Michael R. Solomon, *Advances in Consumer Research,* Vol. 18, Association of Consumer Research, Provo UT, 703-705.

Fredrickson, Barbara L. and Daniel Kahneman (1993), Duration Neglect in Retrospective Evaluations of Affective Episodes, *Journal of Personality and Social Psychology*, 65, 45-55.

Gottman, John M. and Robert W. Levenson (1985), A Valid Procedure for Obtaining Self-report of Affect in Marital Interaction, *Journal of Consulting and Clinical Psychology* , 53, 151-160.

Hughes, David G. (1992), Real-time Response Measures of Television Commercials: Reliability, Construct Validity, New Approaches to Wearout, and Potential Applications for Pretesting, Marketing Science Institute Technical Working Paper no. 92-110.

Kahneman, Daniel (1994), New Challenges to the Rationality Assumption, *Journal of Institutional and Theoretical Economics*, 150, 18-36.

Kahneman, Daniel, 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.

Kahneman, Daniel, and Jackie Snell (1992), Predicting a Changing Taste, *Journal of Behavioral Decision Making*, 5, 187-200

Katz, Karen L., Blair M. Larson and Richard C. Larson (1991), Prescription for the Waiting-in-Line Blues: Entertain, Enlighten, & Engage, *Sloan Management Review*, Winter, 44-53.

Larson, Richard C. (1987), Perspectives on Queues: Social Justice and the Psychology of Queuing, *Operations Research*, 35, 895-905.

Linville, Patricia W., and Gregory W. Fischer (1991), Preferences for Separating and Combining Events, *Journal of Personality and Social Psychology*, 59, 5-21.

Lindley, David (1988), The Waiting Line, *The Washington Post,* November 10th.

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

Nichols, D., E. Smolensky, and T. N. Tideman (1977), Discrimination by Waiting Time in Merit Goods, *The American Economic Review*, 312-323.

Oliver, Richard L. (1980), A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions, *Journal of Marketing Research*, 17, 460-469.

Osuna, Edgar E. (1985), The Psychological Cost of Waiting, J*ournal of Mathematical Psychology*, 29, 82-105.

Redelmeier, Donald A. and Daniel Kahneman (1995), On the Discrepancy Between Experiences and Memories: Patients' Real-Time and Retrospective Evaluations of the Pain of Colonoscopy, Princeton University working paper.

Rothkopf, Michael H., and Paul Rech (1987), Perspectives on Queues: Combining Queues is not Always Beneficial, *Operations Research*, 35, 906-909.

Taylor, Shirley (1994), Waiting for Service: The Relationship Between Delays and Evaluations of Service, *Journal of Marketing*, 58, 56-69.

Tse, David K. and Peter C. Wilton (1988), Models of Consumer Satisfaction Formation: An Extension, *Journal of Marketing Research*, 25, 204-212.

Figure 1: Illustration of a Typical Screen of the Simulation Program

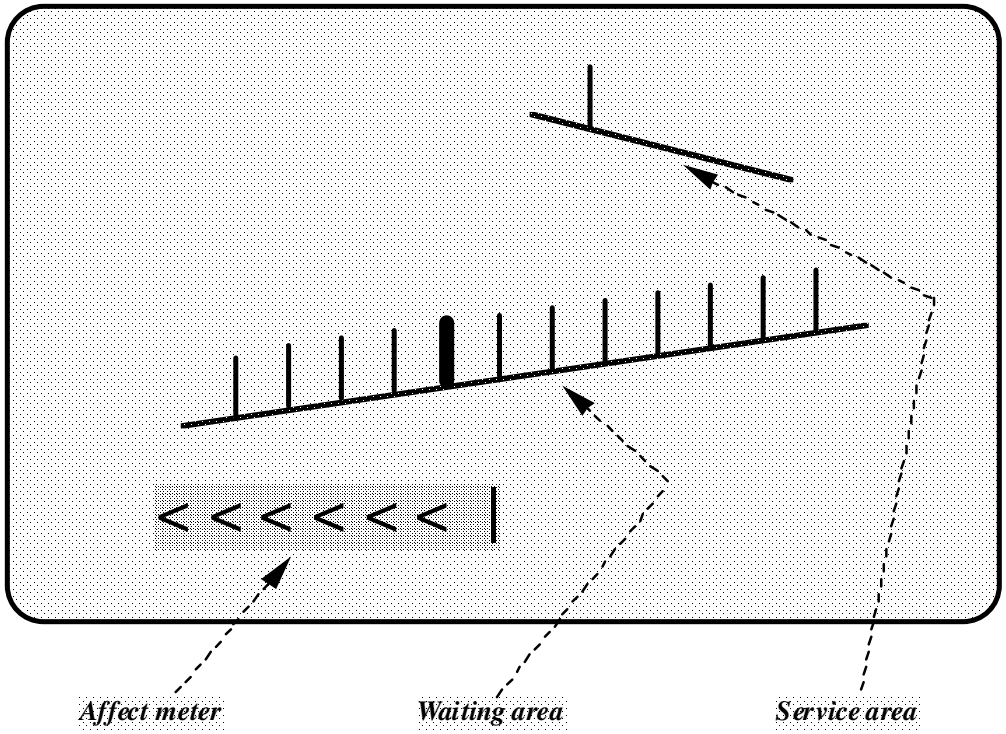


Chart 1: Affective Profiles of Sessions Short, Long & the Three Normal Sessions, Experiment 1A

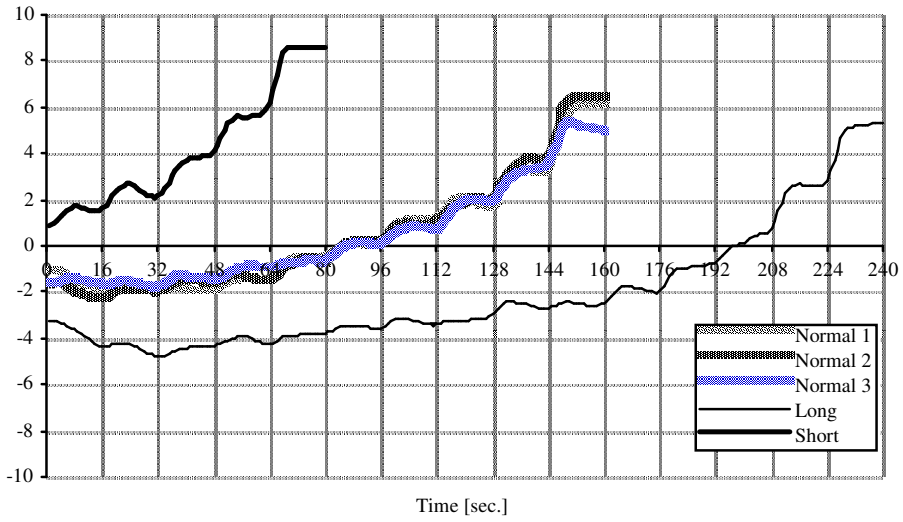


Chart 2: Affective Profiles of Sessions Empty, and Slow, Experiments 1A and 1B

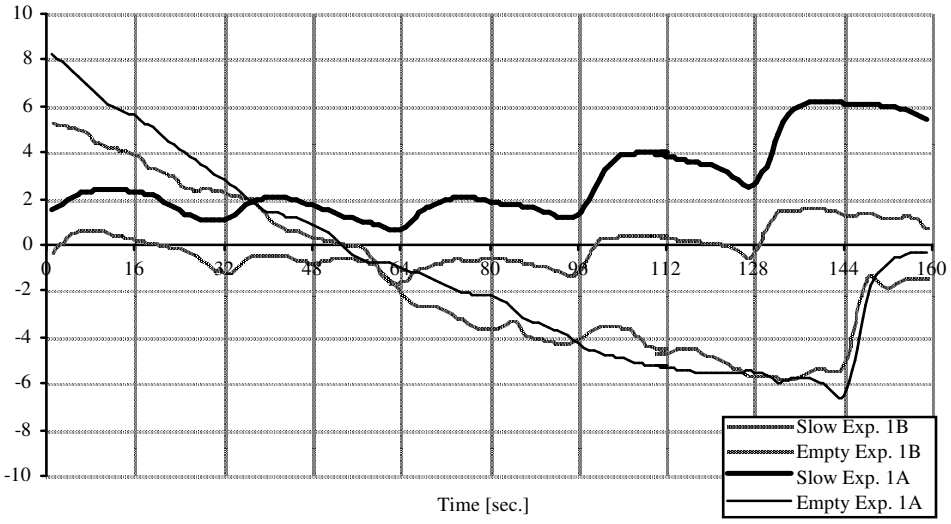


Chart 3: Affective Profiles of Sessions Normal & DOM

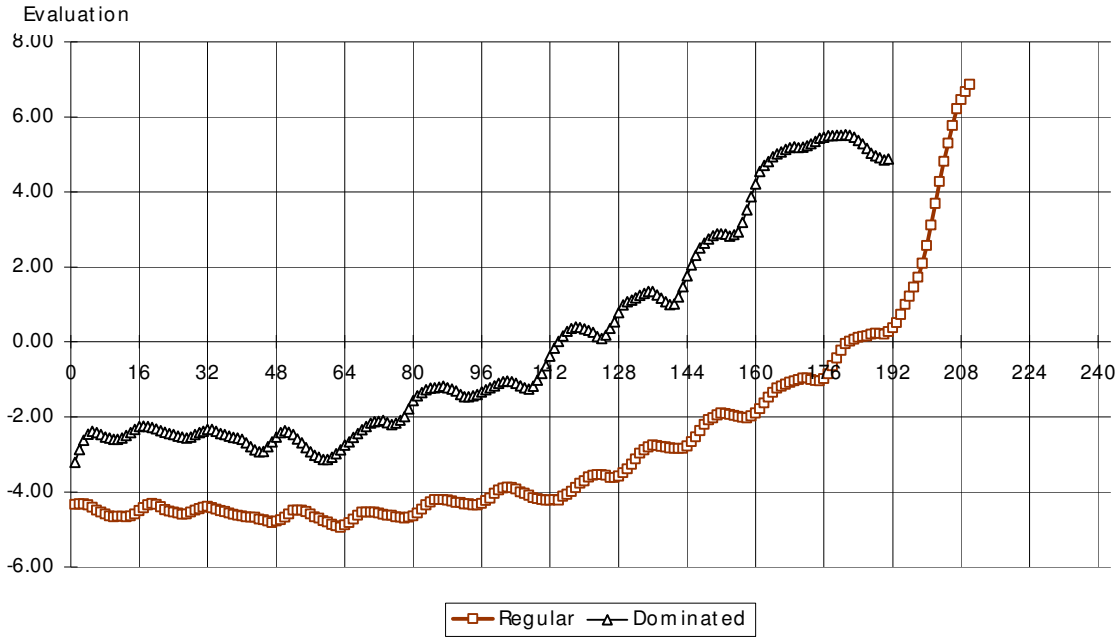
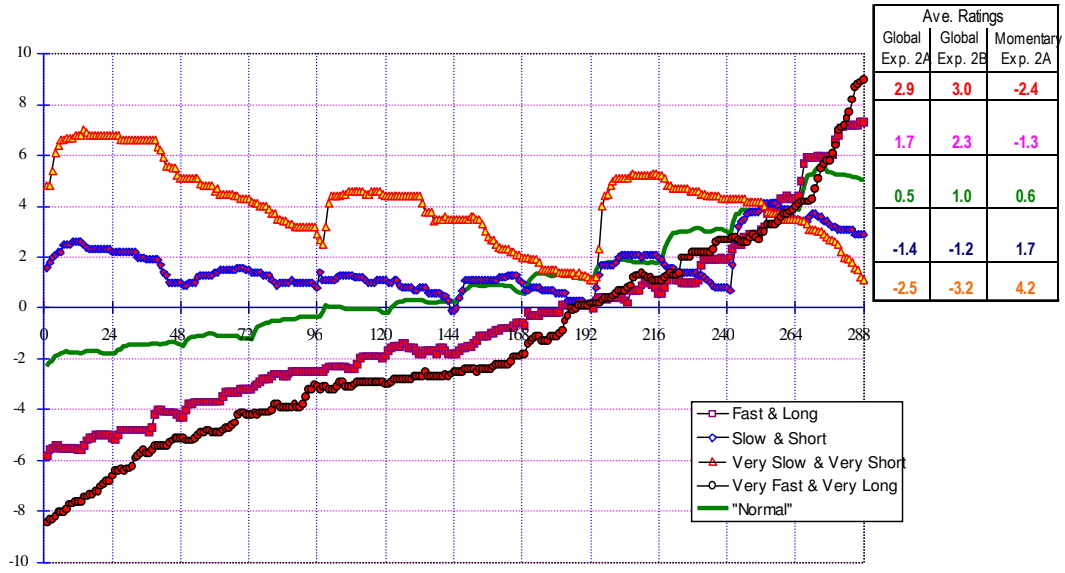


Chart 4: Affective Profiles, Average Momentary Evaluations, & Average Global Evaluations of the 5 Sessions of Experiment 2



Ave. Ratings		
Global Exp. 2A	Global Exp. 2B	Momentary Exp. 2A
2.9	3.0	-2.4
1.7	2.3	-1.3
0.5	1.0	0.6
-1.4	-1.2	1.7
-2.5	-3.2	4.2