On the evaluation of temporally extended experiences

A dissertation presented
by

Irina Cojuharenco

to
The Department of Economics and Business
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of

Management

Universitat Pompeu Fabra
Barcelona, Spain
April 2007

Dipòsit legal: B.42901-2007
ISBN: 978-84-691-1203-8

Thesis advisor Author

Robin M. Hogarth **Irina Cojuharenco**

On the evaluation of temporally extended experiences

Abstract

There are different ways of evaluating experiences lived across time. The "hedometer" paradigm in economics suggests that momentary impressions determine overall evaluations of experiences. There is a mounting empirical evidence for a simple Peak-End rule. Peak (the most extreme) and End (the very last) impressions have been shown to explain overall evaluations without the need to account for the rest.

I test the Peak-End rule using field data from university classrooms and lab data from image-viewing experiments. I find that accounting for average momentary impressions is necessary to explain the variability in overall evaluations of classroom experiences. And yet, End impressions may affect overall evaluations significantly, even after controlling for average momentary impressions. In image-viewing experiments, I study how features of experiences, Peaks, Ends and other, affect post-experience moods. I find that while overall evaluations of experiences are associated significantly with post-experience moods, this is not true for any of the features. Future research should explore additional determinants of overall evaluations that could affect moods.

Using a novel guessing task, I explore lay intuitions about overall evaluations of experiences. The findings suggest that overall evaluations are believed to reflect average momentary impressions. Moreover and alternatively, the personality and attitudes of the experiencing person, experience-specific holistic judgments and behavioral intentions towards experiences are considered to shape overall evaluations.

Finally, in collaboration with Dmitry Ryvkin, I demonstrate why decision-makers may find it difficult to learn the Peak-End rule. Based on field and experimental data, I find that the correlation between Peak-End and average impressions is often high. Thus, learning the Peak-End rule happens under high information redundancy. In a theoretical discussion, I argue that information redundancy depends positively on two factors: (i) the degree of heterogeneity across the experiencing individuals, and (ii) the degree of persistence of momentary impressions within an individual's experience. I show how a model nesting omnipresent psychological processes, that account for these factors, helps explain the magnitude of information redundancy across the data sets analyzed.

Contents

Citations to Previously Published Work

Chapter 4 appeared in the following publication:

> Cojuharenco, I. (2007). Lay intuitions about overall evaluations of experiences, *Judgment and Decision Making*, **2(1)**, 40-47.

Acknowledgments

Completing this doctoral work has been an important experience, and I certainly would not want to forget any of the impressions for the best and the final moments of the way. Robin Hogarth, a person of great wisdom, generosity and tact, advised my work, and it was the greatest privilege of all. Under his guidance, I have discovered a science, Behavioral Decision Making, that continues to motivate my learning, both as a researcher and a person.

I have learnt a lot from Marc Le Menestrel, Rosemarie Nagel and Ton Bosch, and am grateful for their help and encouragement. My work has benefited from fruitful discussions with Jaume Garcia, Albert Satorra, Michael Greenacre, Clare Harries, Brian Gibbs, Daniel Read, Dan Ariely, colleagues at INSEAD and City University Memory Research Unit. During the PhD time and before that, I have met many wonderful teachers, and I thank them.

From my parents and first teachers, Gala and Valeri, I have learnt how much support a loving family can be. More love and support came from Ricardo Nunes, my dearest complice. Friends and colleagues, Merce Roca, Yulia Kasperskaya, Natalia Karelaia, Elena Reutskaja, Irina Yakadina, Elena and Marek Jarocinski, Gisela Bilialova, Carlos Trujillo, Riccardo Pedersini have taught me things from how to think creatively and reason critically to how to best enjoy time spent in Barcelona. I am grateful to the GPEFM Marta Araque, Gemma Burballa, Mireia Monroig and Marta Aragay for support with everything always. Finally, I would like to acknowledge the scholarship of the Spanish Ministry of Education and Science, that made my PhD studies possible and allowed me to visit INSEAD Departments of Marketing and Decision Sciences, where I completed part of this work.

Dedicated to my Teachers

Chapter 1

Introduction and summary

Subject matter and structure of this thesis

Evaluative judgments are fundamental to human decision-making. We pursue activities that we like, dwell on the memories of pleasant events and are willing to repeat an experience, anticipating future liking. Theories of human behavior assume (and sometimes, assume away) processes by which we come to like something, aggregate and tradeoff multiple likes. Economics is unimaginable without the notion of "utility", as psychology is unimaginable without the concept of "pleasure". Everyday decision-making is exposed to the reality of customer satisfaction surveys, job satisfaction reports, performance evaluations, and ultimately, the mundane conversations about how much one adored yesterday's concert, enjoyed today's meal at lunch or liked a seminar presentation.

In this thesis, I explored the nature of the evaluations of experiences lived across time. My research was inspired in an economics "hedometer" paradigm, i.e. the belief that any experience can be represented in terms of satisfaction it gives to the individual experiencing it over time. Then, the overall evaluation of an experience is necessarily a function of momentary impressions throughout the experience's duration. That Peak (most extreme) and End (final) impressions mattered most for overall evaluations was found to be a puzzling empirical regularity. The so-called Peak-End rule puzzled me too. But my research was bound to become interdisciplinary, as I came to believe that the rich psychological reality of experiences lived across time can only be captured partially by the use of hedometers.

This thesis falls naturally into the following three parts:

- Empirical testing of the Peak-End rule in novel environments (Chapter 2) and using novel dependent variables (Chapter 3),
- The study of lay intuitions about overall evaluations of experiences (Chapter 4), and
- The exploration of conditions preventing decision-makers from learning the Peak-End rule (Chapter 5).

The goals and subject matter of the four parts are sufficiently different to merit individual introductions and summaries, which now follow without further ado.

Chapter 2: The Classroom Experience

Research on overall evaluations of temporally extended experiences has found that only maximum and final satisfaction/dissatisfaction with the experience serve as building blocks in its overall evaluation. Field tests of this finding were carried out in hospitals and queues and referred, therefore, to aversive experiences of pain and waiting. I provided a new field test of the finding in a qualitatively different environment of a classroom. In Study 1 undergraduate students reported how engaging they found explanations given to homework problems during a practice session. In Study 2 undergraduate students reported on how interesting they found discussions of a case study. In Study 3 graduate students reported their satisfaction with readings for a readings-based master course. I found that the overall evaluation of these classroom experiences was sensitive to improvements at any and not only certain moments in time.

Chapter 3: Effects of Experiences on Moods

Features of experiences, such as their average momentary affect, peak affect, end affect, variance in affect, affect trend and duration, are known to affect the overall evaluations of experiences. By means of an image-viewing experiment, I test whether effects on post-experience mood valence are similar and mediated by overall evaluations of experiences. Participants viewed either long or short series of either pleasant or unpleasant images, characterized by different patterns of momentary affect. For pleasant experiences, variance in affect had an effect on post-experience mood valence. The effect was mediated by overall evaluations, and model fit almost doubled with the addition of overall evaluations to the predictors of mood valence. For unpleasant experiences, none of the features considered affected post-experience mood valence. Overall evaluations did. These findings suggest that more attention should be devoted to the determinants of overall evaluations beyond features of momentary affect, which contribute significantly to changes in post-experience mood.

Chapter 4: Lay Intuitions about Overall Evaluations of Experiences

Previous research has identified important determinants of overall evaluations for experiences lived across time. By means of a novel guessing task, I studied what decision-makers themselves consider important. As *Informants*, some participants lived and evaluated an experience. As *Guessers*, others had to infer its overall evaluation by asking Informants questions. I rewarded accurate inferences, and analyzed and classified the questions in four experiments involving auditory, gustatory and viewing experiences. Results showed that Guessers thought of overall evaluations as reflecting average momentary impressions. Moreover and alternatively, they tended to consider the personality and attitudes of the experiencing person, experience-specific holistic judgments and behavioral intentions regarding the experience. Thus, according to lay intuitions, overall evaluations reflected more than the experience's momentary impressions.

Chapter 5: Why not learn the Peak-End rule?

Empirical research suggests that evaluations of experiences lived across time are based on the most extreme (Peak) and final (End) impressions, i.e. a Peak-End rule. However, decision-makers believe that evaluations of experiences reflect the average of all impressions without exception. In collaboration with Dmitry Ryvkin, I show why it may be difficult to learn the Peak-End rule. Based on field and experimental data, we find that the correlation between Peak-End and average impressions is often high. Thus, learning the Peak-End rule happens under high information redundancy. In a theoretical discussion, we argue that information redundancy depends positively on two factors: (i) the degree of heterogeneity across the experiencing individuals, and (ii) the degree of persistence of momentary impressions within an individual's experience. We show how a model nesting omnipresent psychological processes of adaptation and anchoring-and-adjustment, that account for these factors, helps explain the magnitude of information redundancy across the data sets analyzed.

Chapter 2

The Classroom Experience

2.1 Introduction

A Teaching Assistant who goes to a class without having had time to prepare, probably hopes that the students would only remember her best explanation. She may have chosen to prepare only the last problem of the problem set hoping to engage students towards the end of the class and so increase the likelihood of a good impression afterwards. In fact, this is exactly the prediction of the research on overall evaluations of experiences. If one assumes that the overall satisfaction/dissatisfaction builds on the satisfaction/dissatisfaction throughout a given experience, empirical research indicates that Peak (the highest) and End (the final) impressions suffice to explain the overall evaluations (Kahneman, 2000). The finding has been termed the Peak-End rule and has undergone testing in a number of laboratory (Varey & Kahneman, 1992; Diener, Wirtz, & Oishi, 2001; Schreiber & Kahneman, 2000; Ariely & Loewenstein, 2000; Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993; Ariely, 1998; Fredrickson & Kahneman, 1993; Baumgartner, Sujan, & Padgett, 1997; Langer, Sarin, & Weber, 2005; Rozin, Rozin, & Goldberg, 2004) and several field studies (Redelmeier & Kahneman, 1996; Ariely & Carmon, 2000; Carmon & Kahneman, 1996). Many laboratory and all field studies involved aversive experiences. There has been a call in the literature for studies involving non-aversive experiences (Fredrickson, 2000; Kahneman, 2000; Ariely & Carmon, 2003).

In what follows, I first review the field studies of the Peak-End rule and present new field studies, involving classroom experiences. A class has a clear beginning and end. Impressions of satisfaction are undoubtedly changing throughout the class duration. The memory of a class is nevertheless only one overall impression. I, therefore, contribute to the literature by testing the Peak-End rule in a qualitatively different field, for which notions of momentary impressions and overall evaluation are natural and appropriate (Ariely & Carmon, 2003).

2.2 Previous field studies

Redelmeier and Kahneman (1996) explored patients' memories of painful medical procedures. They recorded real-time pain reports of patients undergoing colonoscopy ($N = 154$) and lithotripsy ($N = 133$), and subsequently related peak and end pain reports to the reports of overall remembered pain. Real-time pain reports were collected every 60 sec on a 19 cm visual analogue scale. The scale was anchored by statements "No pain" and "Extreme pain" (later, these reports were converted to 0-10 point scale). The overall pain report was done on a 0-10 point scale. Researchers examined the correlation between reports of overall pain and pain throughout the medical procedure. A simple model, including only peak and end reports as predictors of the overall pain, proved adequate to explain the data of the field study. Only small improvement in the accuracy of the prediction was attained by adding total pain, average pain, initial pain and duration of the procedure to the predictors of the overall pain (this comprehensive model explained 69% and 67% of variability in overall pain following colonoscopy and lithotripsy). The simple model explained 67% and 65% accordingly. It performed somewhat worse in explaining the variability of alternative measures of overall pain (e.g. physician's judgment of overall pain). However, the worsening of fit was accompanied by the similar worsening in the prediction of the comprehensive model. The authors of the field study, therefore, interpreted the results as evidence in favor of the Peak-End rule.

Another hospital study (Ariely & Carmon, 2000) took place in a bone marrow transplant unit ($N = 37$). Patients suffering from pains reported pain intensity every hour from 8 a.m. to 6 p.m. on a 0-100 scale. At the end of the day, at 7 p.m. they were asked to report the overall pain. The overall pain had to summarize the day, and patients used the same 0-100 scale. Overall pain ratings were then regressed on the end pain report, peak pain report, the slope of the day's pain profile and average pain report. The regression model was significant with an R-squared of 0.89. The end pain report and the slope of the pain profile proved statistically significant in the prediction of the overall pain, while the average pain rating - did not. The researchers, therefore, concluded that in evaluating experiences people indeed rely on only few select moments and features.

A different result was obtained in a third field study (Carmon & Kahneman, 1996). The overall retrospective evaluation of the aversive experience of queuing was shown to be predicted by only the end report of affect. The result was ascribed to the goal-oriented nature of the experience.

To summarize, all three field studies of the Peak-End rule involved aversive experiences. Both visual analogue scales and ratings were used to elicit subjective experiences of pain/affect. Some experiences were "measured" every minute, others every hour. The same methods were used to test the ability of the Peak-End rule to predict the experience's overall evaluation. Overall evaluations were elicited using ratings. For the sake of comparability, my studies will incorporate these same methods.

2.3 Classroom field studies

I performed 3 distinct classroom field studies. In Study 1 undergraduate students reported how engaging they found explanations given to homework problems during a practice session. In Study 2 undergraduate students reported how interesting they found discussions of a case study. In Study 3 graduate students reported their satisfaction with readings for a readings-based master course. These studies were performed in different academic years using convenience samples.

2.4 Study 1

2.4.1 Method

Task

Participants were asked to evaluate explanations received during a practice session. The instructions stressed that participants were to express their emotional reaction and focus on the evaluation of their affective state and not on any kind of cognitive assessment. Evaluations had to be done on 0-10 point scale anchored by statements “Not engaging at all” and “Extremely engaging”. At the end of the class, they had to evaluate the explanations overall. They used the same 0-10 point scale.

Participants

Participants were undergraduate students aged 19-20, in their first year of either economics or management studies, males and females being represented almost equally. These students attended regular weekly practice sessions of their Economic Theory IV class, for which the author acted as teaching assistant. 36, 42, 46 and 42 students were present in each of the four sessions chosen randomly for the field study. 24 individuals were present in all sessions.

Procedure

In the first practice session, not part of the field study, students were introduced to the task. They were told they would be asked to evaluate how engaging it is to follow explanations given to the problems of the problem set in some practice sessions. They received instructions with an identification letter. I asked them to save the identification letter for later classes. Evaluations had to be done on 0-10 point scales. The word "engaging" in the anchoring statements of the scales was translated into the students' native language to avoid misunderstandings [1], although the language of the class was English. Students signed the agreement to participate, using both their name and the identification letter, and submitted it to the teaching assistant. At this point students did not know about overall evaluations they would have to provide at the end of practice sessions.

At the beginning of every session of the field study students received response sheets. Each response sheet had to be signed by the student using his/her identification letter. Response sheets were named using the name of the problem set of the day. They referred to each point of each problem that had to be covered in class as stipulated in the problem set. Each time there was a horizontal scale from 0 to 10. The rating chosen had to be circled. There was a reminder about the anchoring statements of the scale. Students were told to evaluate each explanation as soon as it finished. I announced the end of each explanation. At the end of the class, response sheets were collected and new forms were distributed for the report of the overall evaluation. Students signed the forms using identification letters (see Appendix A.1).

The last day chosen for the field study students faced a short post-study questionnaire. In this questionnaire, they were asked to situate class experiences on a continuum between activities, such as "going to the dentist" on the extreme left of a visual analogue scale and "watching a pleasant video-clip" on the extreme right. This was done to insure that practice sessions were not perceived as aversive.

[1] "Captivant" was the translation to Catalan.

2.4.2 Results and discussion

75% of students stated that their class experiences were more like “watching a pleasant video-clip”. For each student, we have summarized momentary impressions throughout the class by Peak, End, average rating, the sum of ratings, the slope and the variance of ratings (see Fig 2.1 for a sample chart of momentary impressions). Descriptive statistics of these variables and the overall evaluations of practice sessions are presented in Table 2.1.

Table 2.2, column 1, states the results of a standard regression of overall evaluations on the features of momentary impressions. We have used pooled ordinary least squares after verifying the non-significance of differences in coefficient estimates in fixed versus random effects panel data analysis (Hausman test $\chi^2(6) = 4.65$, *ns*). Regression output shows a large and significant effect of average rating and a five times smaller, but significant effect of the End rating on the experience's overall evaluation. The significance of End rating confirms the possibility for “impression management” throughout the class. If the teaching assistant decided to invest into making one point of the problem set particularly engaging, her effort should be directed at the last point in the problem set. The contribution of any improvement anywhere in the problem set would be reflected in the average rating. However, if improvement concerned the End point, the effect on the overall evaluation would be maximized.

The importance of average real-time rating was not expected, given the results of previous studies. Some elements of the procedure employed in Study 1 could be the cause. In particular, the format of reports for momentary impressions and the overall evaluation was uniform, and students reported momentary impressions on a single

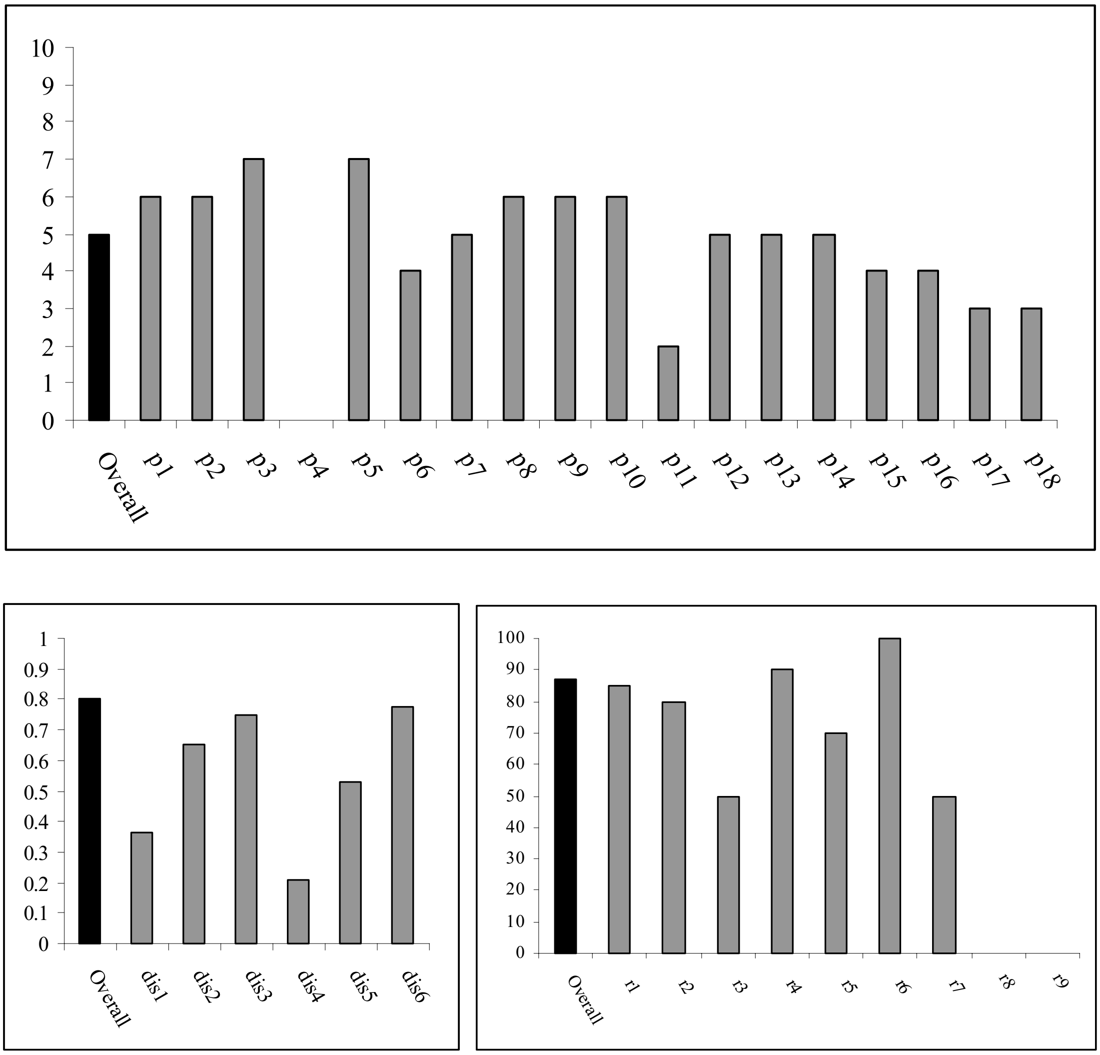


Figure 2.1: Charts of sample class experiences. Upper panel: Study 1 participant evaluating class explanations; lower panel left: Study 2 participant evaluating class discussions; lower panel right: Study 3 participant evaluating readings done for the course.

Table 2.1: Descriptive statistics of overall evaluations and ratings of momentary impressions.

	Mean	Std. dev.	Minimum	Maximum
Study 1 ($N = 165$)				
Overall Evaluation	6.7	1.2	4	10
Average Rating	6.6	1.2	3	10
Sum of Ratings	56.7	30.8	16	142
Peak Rating	8.1	1.1	5	10
End Rating	6.7	1.9	0	10
Ratings Variance	2.0	2.2	0	13.5
Ratings Slope	.1	.4	-.9	1.9
Study 2 ($N = 28$)				
Overall Evaluation	.68	.13	.40	.90
Average Rating	.59	.13	.36	.85
Peak Rating	.81	.11	.65	1.00
End Rating	.59	.21	.14	.93
Ratings Variance	.03	.03	.01	.12
Ratings Slope	.00	.03	-.05	.10
Study 3 ($N = 17$)				
Overall Evaluation (points/bar)	80/78	9/13	65/54	95/94
Average Rating	77	9	59	89
Sum of Ratings	561	136	356	765
Peak Rating	89	8	70	100
End Rating	76	17	30	95
Ratings Variance	115	128	12	458
Ratings Slope	1	3	-5	6

response sheet. They could easily see their previous reports which looked like a set of lines bounded by a circled value on rating scales. There is evidence that physical sets of objects are instantaneously perceived in terms of their average characteristic, e.g. a set of lines in terms of their average length (Treisman & Chong, 2003). This could favor the predictive ability of average momentary impressions. The procedure for eliciting momentary impressions was changed in the next field study.

2.5 Study 2

2.5.1 Method

Task

Participants were asked to evaluate discussions along the questions of a weekly case study. There were 6 issues to be discussed. Participants were asked to evaluate each discussion "just to see how we enjoy what we do in this class" and assured of anonymity. After each discussion they had to complete the statement "I liked the discussion we had" by marking a single vertical slash through a 72 mm visual analogue scale, anchored at the left with "Not at all" and on the right with "Very much". At the end of the class, they answered additionally "How much did you enjoy all class discussions today?" using a 0-100 point scale. "0" meant the student did not like the discussions at all and "100" meant the student liked the discussions very much.

Table 2.2: Overall evaluations of class experiences explained by the features of momentary impressions.

Dependent Variable	Overall Evaluations				
	Class Explanations	Class Discussions	Class Readings		
			points	bar	2-item scale
	(1)	(2)	(3)	(4)	(5)
	Coefficient (t-statistic)	Coefficient (t-statistic)	Coefficient (t-statistic)		
Predictors					
Average Rating	0.743	1.393	-1.053	-1.409	-0.115
	(6.87)**	(4.11)**	(1.37)	(0.92)	(1.14)
Sum of Ratings	-0.003		0.054	0.050	0.005
	(1.49)		(2.16)	(1.00)	(1.52)
Peak Rating	-0.081	-0.537	1.808	2.628	0.205
	(0.84)	(1.71)	(2.07)	(1.52)	(1.79)
End Rating	0.143	0.030	-0.406	-0.965	-0.061
	(2.67)**	(0.20)	(0.99)	(1.19)	(1.14)
Ratings Variance	0.023	3.111	-0.078	-0.151	-0.010
	(0.60)	(1.90)	(1.22)	(1.18)	(1.23)
Ratings Slope	-0.168	-0.889	1.247	1.447	0.127
	(0.91)	(1.04)	(0.97)	(0.57)	(0.75)
Constant	1.685	0.179	8.983	14.521	-6.473
	(4.37)**	(1.53)	(0.45)	(0.37)	(2.48)*
Observations	165	28	17	17	17
R-squared	0.71	0.73	0.74	0.49	0.62

* significant at 5%; ** significant at 1%

Participants

Participants were 28 undergraduate students aged 18-20, in their first year of management studies. There were 19 female and 9 male students. These students attended regular weekly practice sessions of their Business Economics II class, for which the author acted as teaching assistant.

Procedure

The field study was carried out during a practice session of Business Economics II. Before class discussions started, every student received an empty blank envelope and seven stapled response forms turned backside up. Six forms contained the six case study questions in the order in which they were subjected to discussion. Next to the discussion question in each of the six forms, there was the statement "I liked the discussion we had" followed by the visual analogue scale. Every time a discussion ended the teaching assistant invited the students to detach the response form from the stapled pack, turn it face up, fill it in, and place it into the envelope. The seventh form was turned face up at the very end of the practice session, after the last discussion had been evaluated. It asked for the overall evaluation of the discussions. The overall evaluation had to be written on the envelope, which, by that time, contained evaluations for the six discussions (see Appendix A.2). Envelopes were left on a specially indicated chair as students left the room.

2.5.2 Results and discussion

Table 2.2, column 2, summarizes results obtained in Study 2. In the standard regression of overall evaluation on the features of momentary impressions in the classroom, only average real-time rating turned out significant at 5% level. Ratings variance was significant at 10% level. These results suggest that overall evaluations did not build on only few representative moments of the experience. Parallels between Study 2 and Study 1 will be discussed in section *Results summary*.

2.6 Study 3

2.6.1 Method

Task

Participants were asked to report satisfaction with the weekly readings assigned for a trimester-long master course in Behavioral Decision Making. They used a 0-100 scale anchored by the statements "Did not like it at all" and "Enjoyed it very much". At the end of the course, they were asked to evaluate the readings overall. They used both a visual analogue scale (further referred to as *bar data*) and assigned points on a 0-100 scale. The extremes of both instruments were anchored by the statements "Did not like it at all" and "Enjoyed it very much".

Participants

Participants were 22 graduate students enrolled in a readings-based master course. For reasons of anonymity I could not collect information about the age and the sex of participants.

Procedure

Graduate students enrolled in the course had to read 2 to 4 academic papers weekly. They had to select one paper for presentation during a class, and all classes of the course were devoted to the discussion of papers assigned for the corresponding week. Classes were held on Mondays and Tuesdays, and there were 9 reading assignments in total. Every Tuesday students were asked to report their satisfaction with the reading they had done for that week. They were given a response form in which they had to identify themselves by initial letters of their first and last names. The readings assigned for the week were listed in a table. The student had to mark readings he/she had done, and evaluate each in points, using a 0-100 scale. If students had read more than a single paper, or papers assigned for other weeks, they had to write down the titles of papers and provide a global evaluation of the readings done. The last Tuesday of the trimester students had to fill out the usual response form with an additional question about all the readings done for the course (see Appendix A.3).

2.6.2 Results and discussion

Study 3 results are presented in Table 2.2, columns 3-5. Evident are the problems of sample size in this study, which has been reduced to 17 because 5 students did not come to class the last day of the study. The sum of ratings and the rating of the favorite week turn out significant in predicting the overall evaluation when the evaluation is done in points at 10% significance level. Curiously, for 71% of students the week of readings that was evaluated the highest was the week during which the student had to make his/her presentation. None of the predictors used exerted a significant partial effect on overall evaluations as measured using the horizontal bar. I used Chronbach's alpha to create a composite measure of overall evaluation using bar data and points data. Reliability of this 2-item scale was 0.92 [2], however its use did not clarify the results.

2.7 Results summary

I performed classroom field studies in order to test the Peak-End rule. Drawing conclusions based on the results in Table 2.2 requires important assumptions. I have chosen to revert to a simpler analysis. In particular, I explored the zero-order correlation of average momentary impressions with overall evaluations in all experiences studied. I compared it to the zero-order correlation of the Peak-End rule (or Peak-only rule, whichever performed best) with the same overall evaluations, and tested the significance of differences in these correlations, using Hotelling's t-test. Moreover, I have compared the mean absolute errors of predictions of the overall evaluations based on average impressions versus the Peak-End rule. Results are presented in Table 2.3.

In Study 1, the prediction by average momentary impressions is clearly superior in both its correlation with actual overall evaluations (Hotelling $t(162)$=3.85, $p < 0.01$) and level fit (M=.85, SD=.06 for the mean absolute error of the Peak-End rule, M=0.50, SD=0.04 of average, $t(164)$=5.51, $p < 0.01$). In Study 2, the prediction by

[2] Average interitem correlation was 0.86.

Table 2.3: Peak-End rule and Average momentary impressions as predictors of overall evaluations of classroom experiences.

Average rating	Scale of overall evaluations	Correlation with overall evaluations	Mean abs. error
Study 1 ($N = 165$)	0 to 10	.83	.50
Study 2 ($N = 28$)	0 to 1	.82	.11
Study 3 ($N = 17$)			
– points	0 to 100	.63	6
– bar	0 to 100	.44	9
– 2-item scale	–	.54	–

Peak-End rule	Scale of overall evaluations	Correlation with overall evaluations	Mean abs. error
Study 1 (N=165)	0 to 10	.73	.85
Study 2 (N=28)	0 to 1	.51	.09
Study 3 (N=17)			
– points	0 to 100	.76	8
– bar	0 to 100	.60	11
– 2-item scale	–	.69	–

average momentary impressions is superior in its correlation with overall evaluations (Hotelling $t(25)$=-3.34, $p < 0.01$) and equivalent in terms of fit with the prediction by Peak-End rule (M=.09, SD=.02 for the mean absolute error of the Peak-End rule, M=0.11, SD=0.01 of average, $t(27)$=-.73, *ns*). In Study 3, there are no significant differences.

Thus, average momentary impressions played an important role for overall evaluations of the classroom experiences. Even though I am not able to affirm that "impression management" is impossible altogether, I have strong grounds to claim that the improvement at any point during the experience would be reflected in its overall evaluation.

2.8 General discussion

I tested the Peak-End rule in the classroom. Classroom experiences chosen for field studies were different. In Study 1, participants were relatively passive, while in Study 2, it was participants' interventions during class discussions that "created" the experience. How these experiences compare to pain and waiting, or similar experiences, is not yet clear. The generalization of findings from these and other fields will be possible as more field studies are undertaken. Although Study 3 results were inconclusive, in Study 1 and Study 2, average momentary impressions provided a good prediction about both the level of overall evaluations for individual participants and differences across participants. Peak and End impressions were considerably worse in predicting the differences. Thus, overall evaluations of classroom experiences have been shown to be less selective than expected.

There may be two potential explanations for this lack of selectivity. First, classroom experiences may not have been emotional enough. Langer et al. (2005) found that people evaluated sequences of monetary payments quite objectively when these were received without affection, but evaluated the same sequences according to the Peak-End rule when payments were linked to effort and evoked a stronger affective reaction. Second, classroom experiences may not have been fluid enough. Ariely and Zauberman (2000, 2003) investigated the effects of breaking and combining experiences on their overall evaluation, and arrived at the conclusion that the more fluid segments were evaluated by the Peak-End rule. Segment-specific evaluations were then combined in a more equal weighting fashion to produce the overall evaluation. The emotional character of a classroom may be difficult to alter. The present findings are informative to the extent to which a relative coldness and discreteness are inherent in a given classroom experience.

Chapter 3

Effects of Experiences on Moods

3.1 Introduction

If you consider a particular movie to have been a pleasant experience, does it mean you walk away from the movie theater in a happy mood? Better endings, scenes provoking feelings of a dramatically different intensity, duration of one hour versus three hours are some of the determinants of more favorable overall evaluations of the movie. In fact, the effects of such features of experiences on overall evaluations have been studied extensively (Kahneman, 2000). However, the effects on mood states have not been investigated. And yet, post-experience mood may be determining largely how you will relate to other people, whether you will or will not scrutinize advertisement messages on the way back, or how many transportation options you will consider to go home that night.

Unlike overall evaluations of experiences, moods do not have specific referents. Mood is a general, diffuse feeling state, typically operationalized as the individual's response to the question "How do you feel right now?" Moods are known to affect everyday decisions and behavior, influence memory, attitudes towards other people, the information we choose to consider in a decision-making context, our motivation to act, etc. (Schwarz & Clore, 2006; Forgas, 1995; Wegener, Petty, & Smith, 1995). Effects of moods have been tested in psychological laboratories and discussed widely in literatures from organizational to consumer behavior (Robbins & DeNisi, 1998; Forgas & George, 2001; Gardner, 1985; Bagozzi, Gopinath, & Nyer, 1999). Good and depressed moods have been induced by exposing participants to affect-laden texts, images, movies, by giving them small unexpected gifts or asking the decision-makers to recall happy or sad life episodes (Luomala & Laaksonen, 2000). In doing so, researchers have endorsed the view that moods are both background states and responses to previous experiences that affect future behavior and experiences (Gnoth, Zins, Lengmueller, & Boshoff, 1999). Definitions of mood that refer to it as a *remainder* of a strong emotion (Pieters & Van Raaij, 1988) support that view. And yet, there is much to learn about the effects of experiences on moods. If patterns of momentary affect in experiences can be controlled, e.g. by choosing to show a pleasing scene towards the end of the movie, one could design experiences so as to affect moods in predictable ways. Such effects of experiences on moods could be mediated by overall evaluations of experiences or affect them over and beyond overall evaluations.

I investigate the link between overall evaluations of experiences, patterns of momentary affect in them, and moods induced. In the following section, I define the terminology and explicate my research question. Next, I review the literature on the evaluation of experiences in order to identify features of experiences to be considered. I then describe the design of the image-viewing experiment used to explore the effects of experiences on moods. I conclude by discussing the findings.

3.2 Terminology and research question

Throughout this work, I will use the following terminology. The term “momentary affect” will refer to an individual’s emotional reaction at the moment of exposure to a given affective stimulus, e.g. the perceived pleasantness/unpleasantness of an affective image when looking at it. The term “overall evaluation” will refer to the remembered pleasantness/unpleasantness of the exposure to all the affective stimuli, constituting the experience, e.g. the overall pleasantness/unpleasantness of 30 images, recalled after viewing them. Finally, the term “mood” will refer to an individual’s subjectively perceived referent-free post-experience affective state, e.g. feeling happy or feeling depressed.

I hypothesize that features of the experience that characterize the evolution of its momentary affect determine the overall evaluation of the experience (arrow A in Figure 3.1). Overall evaluation, then, affects post-experience mood (arrow B). Thus, features of the experience can be shown to exert an indirect effect on mood (arrow C). The effect is called indirect because it is mediated by the overall evaluation. The effect of the features of the experience on mood is expected to have the same sign as their effect on the overall evaluation.

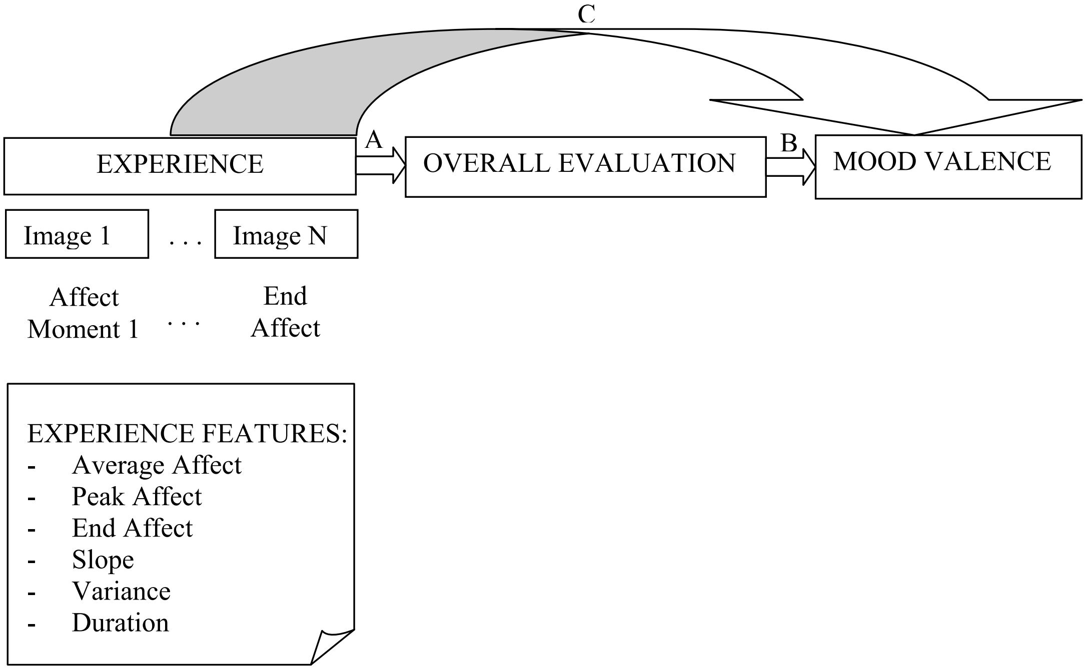


Figure 3.1: Hypothesized effects of experiences on mood valence.

3.3 Features of experiences

A significant empirical effort has been devoted recently to the study of the features of experiences that determine their overall evaluations in retrospect (Kahneman, 2000). I discuss the most commonly considered features below.

3.3.1 Peak affect

Daniel Kahneman and his colleagues were the first to demonstrate that people gave a disproportionately high weight to Peak affect intensity in evaluating experiences lived across time. Typically, in the experiments they conducted participants were asked to report how much they liked/disliked certain affective stimuli at the moment of exposure, and later how much they liked/disliked the experience overall. Fredrickson and Kahneman (1993) conducted an early experiment in which participants viewed aversive film clips and pleasant film clips that varied in duration and intensity. Participants provided real-time ratings of affect during each clip and overall evaluations of each clip when it was over. There were short and long clips [1]. It was found that overall evaluations did not track average momentary affect or total affect (the sum of momentary affect). Rather, regression analysis showed that the average of only Peak and End (final moment) affect was the best predictor of overall evaluations. Fredrickson (2000) argued that Peaks (and Ends) in both pleasant and unpleasant experiences carried information about a person's coping capacity, were worth remembering, and therefore, overall evaluations depended positively on peak affect.

3.3.2 End affect

The importance of End affect for the evaluation of experiences lived across time could be attributed naturally to *recency* effects (Hogarth & Einhorn, 1992). And yet, Fredrickson (2000) has shown that Ends matter for a different reason. She created short-term social relationships in a laboratory setting. After each session, each person viewed a videotaped portion of their conversation and provided moment-by-moment ratings of how they were feeling during the actual interaction. Half the pairs believed that the new acquaintanceship would end the same day it started, and half thought

[1] On average, short pleasant clips lasted for 37 sec and long for 111 sec. Short aversive clips lasted for 29 sec and long for 84 sec.

it would continue on a subsequent day. It was shown that endings were weighted heavily in people's global evaluations of the experience, but not when the experience was expected to continue. Although equally recent, known endings produced effects that were different from the effects of endings that were unknown. Recency could not explain this finding.

3.3.3 Average affect

Overall evaluations are often assumed to reflect average momentary affect. A number of empirical studies that tested the importance of Peaks and Ends for overall evaluations of experiences represented attempts to revisit this assumption (Baumgartner et al., 1997; Stone, Shiffman, & DeVries, 1999). The Merriam-Webster's online dictionary defines the word "overall" in terms of "general", and "general" in terms of "prevalent" and "average". This is significant, as dictionaries are meant to systematize the way concepts are used in everyday life. Although the perceived equivalence between "overall" and "average" appears to be inconsistent with the importance that Peaks and Ends were found to have for overall evaluations, there are boundary conditions to the Peak-End phenomena worth mentioning. Partitioning of experiences, explicit momentary ratings and low emotionality of the stimuli have been shown to tilt the balance in favor of average momentary affect as a predictor of overall evaluations (Ariely & Zauberman, 2000, 2003; Ariely, 1998; Langer et al., 2005).

3.3.4 Trend in affect

A significant body of research has shown that trends in affect are important determinants of preferences over sequences of future outcomes. People prefer experiences with improving trends. Experiences that are increasingly pleasant or less unpleasant are preferred to experiences that decrease in pleasantness or are increasingly unpleasant (Loewenstein, 1987; Hsee & Abelson, 1991). Ariely and Carmon (2000) showed additionally that trends matter similarly for overall evaluations. They conducted a hospital field study. Once every hour (between 8 a.m. and 6 p.m.) a nurse asked participating patients to rate the pain they experienced on a 0–100 scale (0 represented no pain, and 100 the worst pain they could imagine). At the end of the day (7 p.m.) the nurse asked each patient to rate the overall pain they experienced throughout that day on the same 0–100 scale. The rate of change in the hour-by-hour pain ratings turned out to be a significant predictor of the overall pain evaluations. Upward-sloping trends in hourly pain were associated with greater overall pain evaluations.

3.3.5 Variance in momentary affect

Variance in momentary affect, further referred to as *affect variance*, is another feature of experiences that could be important for their overall evaluations. People have been shown to prefer variability in sequences of future outcomes and to spread multiple desirable or undesirable experiences out over time (Read, Loewenstein, & Rabin, 1999; Read, Antonides, Ouden, & Trienekens, 2001; Loewenstein & Prelec, 1993). Implications of these findings for the effect of affect variance on overall evaluations have not been explored.

3.3.6 Experience duration

That duration of experiences should matter for their evaluation has been considered a compelling logical principle (Fredrickson & Kahneman, 1993; Kahneman, Wakker, & Sarin, 1997). Few people would want additional moments of pain, even if mild, and many should want to prolong pleasant experiences, even if additional moments would be only moderately pleasant. Thus, greater duration was believed to lead to more intense overall evaluations. And yet, overall evaluations of experiences have not been found to depend on the duration of experiences as believed. The multiplicity of mood-induction procedures, some lengthy and others short, suggests that moods, too, may be duration-insensitive.

3.4 Method

3.4.1 Design

Experimental subjects were randomly assigned to one of the four image-viewing experiences: the viewing of a long pleasant series (30 images, viewing time 4.5 min, treatment L+), long unpleasant series (30 images, viewing time 4.5 min, treatment L–), short pleasant (3 images, viewing time 27 sec, treatment S+) or short unpleasant (3 images, viewing time 27 sec, treatment S–). Long series of images contained the short series and 27 additional images. A post-experimental questionnaire, containing the measurement of mood, was administered after the viewing experience in each treatment.

3.4.2 Stimuli

The International Affective Picture System, IAPS, (Lang, Bradley, & Cuthbert, 2005) was the main source of images used in the experiment. A few images were taken from the pre-tests of a set collected by the author through the World Wide Web. Image series were compiled using the valence, arousal and dominance ratings in the IAPS. Pleasant series consisted of images portraying happy people, beautiful landscapes and scenes of animal life, as well as neutral displays of objects, such as a mug or a plate. Unpleasant series consisted of images showing repulsive insects, dirt, traffic accidents, homicide, and sad scenes of death and despair.

In order to separate the effects of experience duration, the intention was for short and long experiences (i.e., experiences in treatments L+ and S+, and experiences in treatments L– and S–) not to differ in terms of average momentary affect, peak and end affect and affect slope (given equal average and peak affect, affect variance would be naturally lower in longer experiences). There were four different orders of images in every experimental treatment. The intention was to construct series in which affect, positive or negative, would increase, decrease or change in a hill- or a valley-like fashion throughout the viewing. Images with serial positions 1, 15 and 30 in long experiences constituted the short series analogues for each of these patterns. Certain variability in patterns of affect throughout experiences was, thus, insured (see Figure 3.2). Although different in valence intensity, images were chosen so as to be comparable on the dimensions of arousal and dominance.

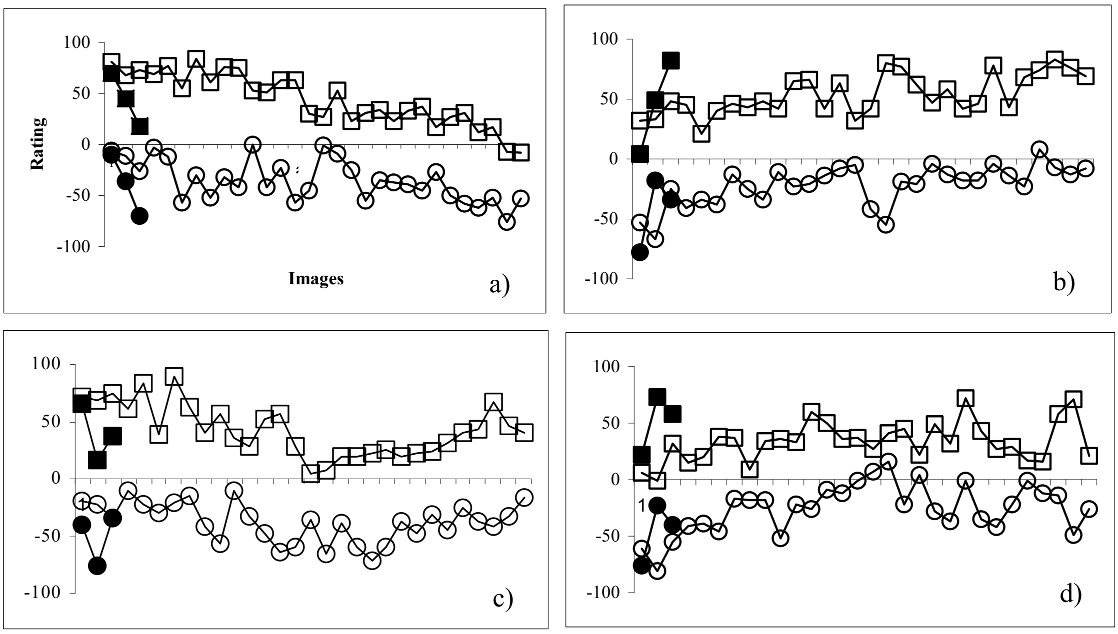


Figure 3.2: Patterns of momentary affect in pleasant and unpleasant experiences (worsening pattern of image evaluations in a), improving pattern in b), valley-like and hill-like patterns in c) and d)).

3.4.3 Task

In all treatments, participants evaluated the pleasantness/unpleasantness of each image viewed on a scale, ranging from -100 to +100. The portion of the scale from -100 to 0 had to be used for images from very unpleasant to neutral, and the portion from 0 to +100 for images from neutral to very pleasant. In addition, they were asked to fill out a post-experimental questionnaire (see Appendix B). Questionnaire items pertinent to the present study included (i) the evaluation of the experience lived on a horizontal bar anchored by statements "This was an absolutely horrible experience" on the left, "This was a neutral experience" in the center, and "I experienced a great deal of pleasure" on the right [I will further refer to this measure of overall evaluation as *O-bar*]; (ii) the evaluation of the experience in points on a scale ranging from -100 ("disliked it completely") to +100 ("liked it very much") [further referred to as *O-points*]; (iii) the evaluation of the series viewed in points on a scale ranging from -100 ("the series was totally unpleasant") to +100 ("the series was totally pleasant") [further referred to as *O-serie*]; (iv) the description of the image that would be the most representative of the overall impression from the series; (v) current mood report using the 9-point scale Self-Assessment Manikin (Bradley & Lang, 1994) [2] ; (vi) the estimate of mean image rating; and (vii) the recall of the contents for the first, last, most pleasant/unpleasant, and least pleasant/unpleasant images.

[2] The Self-Assessment Manikin (SAM) is a non-verbal pictorial assessment technique that directly measures the valence, arousal, and dominance associated with a person's mood. Each dimension is portrayed with a graphic character arrayed along a continuous 9-point scale. For valence, SAM ranges from a happy, smiling figure (9 points) to an unhappy, frowning figure (1 point). For arousal, SAM ranges from sleepy, with eyes close (1 point), to excited, with eyes open (9 points). For dominance, SAM ranges from a very small figure, representing the feeling of being controlled or submissive (1 point), to a large figure, representing in-control, or a powerful feeling (9 points).

3.4.4 Participants

There were 89 participants, recruited through the web-based system of the experimental economics laboratory at Universitat Pompeu Fabra. 32 were male, average age was 21. All were undergraduates, and the vast majority were students in economics. 23 participants viewed the long pleasant series, 23 the long unpleasant, 21 the short pleasant and 20 the short unpleasant series. 2 female participants were excluded from the analysis as they gave positive evaluations to unpleasant images. Participants received a fixed participation fee (7 euros in treatments L+ and L–, and 5 euros in S+ and S–).

3.4.5 Procedure

Participants entered a computer laboratory, sat in isolated workstations, and the experimenter read aloud a general description of the experiment. The experiment was described as a study of how people respond to images representing various life events. Participants were told that images would be presented to them in a PowerPoint presentation, and remain on the screen for 9 seconds for evaluation. Evaluations had to be done in writing, using evaluation sheets to be found inside envelopes on their tables. Each evaluation sheet provided the space for 2 image evaluations. Images were referred to by their serial position, e.g. "Image 1" for the first image of the series. Images could range from very unpleasant to very pleasant, and everybody was asked to sign a consent form in order to participate. A PowerPoint presentation was open in every workstation and participants were invited to advance it and read through the first two slides of instructions. The first slide familiarized participants with the scale to be used in evaluating image pleasantness/unpleasantness. The second slide contained 3 miniature sample affective images: a burn victim on the extreme left of the screen, a chair in the middle of the screen, and a beautiful beach on the extreme right of the screen. Participants were informed about the number of images to be evaluated and asked to advance the presentation in order to start the viewing. Once participants had evaluated all images, they were asked to fill out the post-experimental questionnaire, after which they were thanked, paid and dismissed.

3.5 Results and discussion

3.5.1 Experiences

Measures *O-points*, *O-bar*, and *O-series* were used to construct a 3-item scale measuring overall evaluations of experiences (standardized items). Scale reliability coefficient was 0.97 as measured by Chronbach's alpha [3]. Another measure of overall evaluation employed followed literally Milan Kundera's metaphor (in his novel Immortality) "memory does not make films, it makes photographs", often cited in the literature on overall evaluations of experiences. Participants described the snapshot of the image-viewing experience, i.e. the image that best represented their overall impression from the series. I used the rating of the image they had indicated as an alternative measure of overall evaluation [Representative Rating].

I summarize viewing experiences in each experimental treatment by their overall evaluations (3-item scale and the representative rating), average image evaluation

[3] Average interitem correlation was 0.93.

Table 3.1: Pleasant experiences: descriptive statistics of overall evaluations, features of experiences and mood valence.

	Mean	Std. dev.	Minimum	Maximum
Long experiences (L+, N=23)				
Overall (3-item scale)	.99	.54	-.24	1.78
Rep. Rating	72	26	0	100
Mood Valence	7.30	1.14	5	9
Average Affect	44	18	15	79
Peak Affect	90	12	50	100
End Affect	31	39	-50	100
Slope	-.6	1.8	-4.4	2.3
Variance	833	420	179	2042
Short experiences (S+, N=21)				
Overall (3-item scale)	.66	.52	-.24	1.66
Rep. Rating	59	35	0	100
Mood Valence	6.95	1.20	5	9
Average Affect	45	19	0	83
Peak Affect	75	26	0	100
End Affect	52	36	0	100
Slope	7.9	29.6	-50	45
Variance	1246	817	0	2908

[Average Affect], the most positive/negative image evaluation in pleasant/unpleasant experiences respectively [Peak Affect], the evaluation of the last image in the series [End Affect], the slope of a linear trend in image evaluations [Slope], and the variance of image evaluations [Variance]. See Table 3.1–3.2 for descriptive statistics.

The intention for long and short experiences not to differ in features other than affect variance and duration was not reflected by the data. Momentary affect in long pleasant experiences was characterized by a higher peak (M=90, SD=12 in L+, M=75, SD=26 in S+), $t(28)$=2.42, $p < 0.05$, a lower end (M=31, SD=39 in L+, M=52, SD=36 in S+), $t(42)$=-1.85, $p < 0.05$, and a lower variance (M=833, SD=420

Table 3.2: Unpleasant experiences: descriptive statistics of overall evaluations, features of experiences and mood valence.

	Mean	Std. dev.	Minimum	Maximum
Long experiences (L–, N=23)				
Overall (3-item scale)	-.84	.32	-1.39	-.31
Rep. Rating	-57	35	-100	10
Mood Valence	3.83	1.53	1	7
Average Affect	-30	14	-76	-12
Peak Affect	-92	9	-100	-80
End Affect	-26	35	-100	10
Slope	.03	1.28	-2.23	1.91
Variance	833	329	375	1472
Short experiences(S–, N=20)				
Overall (3-item scale)	-.88	.46	-1.92	-.24
Rep. Rating	-62	36	-100	0
Mood Valence	4.75	1.48	2	9
Average Affect	-45	21	-83	0
Peak Affect	-80	16	-100	-50
End Affect	-43	34	-100	0
Slope	4.88	27	-50	45
Variance	1250	810	133	2858

in L+, M=1246, SD=817 in S+), $t(29)$=-2.08, $p < 0.05$. There were no significant differences in terms of average affect or affect slope. Overall, long pleasant experiences were remembered as significantly more pleasant than short pleasant experiences (M=0.99, SD=0.54 in L+, M=0.66, SD=0.52 in S+), $t(42)$=2.02, $p < 0.05$ [4].

Momentary affect in long unpleasant experiences was characterized by a higher average (M=-30, SD=14 in L–, M=-45, SD=21 in S–), $t(32)$=2.71, $p < 0.01$, a more negative peak (M=-92, SD=9 in L–, M=-80, SD=16 in S–), $t(29)$=-2.97, $p < 0.01$, a more positive end (M=-26, SD=35 in L–, M=-43, SD=34 in S–), $t(41)$=1.61, $p < 0.10$, and a lower variance (M=833, SD=329 in L–, M=1250, SD=810 in S–), $t(24)$=-2.15, $p < 0.05$. There were no significant differences in terms of affect slope. Overall, long and short unpleasant experiences were remembered as comparably unpleasant [5].

3.5.2 Effects of experiences on moods

Experimental treatments altered the mood of participants. The distribution of valence ratings is shown in Figure 3.3. Intensity of arousal and control were not affected [6]. In terms of mean valence ratings, participants in treatment L– rated their mood the lowest, followed by those in treatment S– (M=3.82, SD=1.53 in L–, M=4.75, SD=1.48 in S–), $t(41)$=2.02, $p < 0.05$. Those in treatment L+ rated their mood valence most positively, followed by those in treatment S+, although here the difference in mean valence ratings was not statistically significant (M=7.30, SD=1.14 in L+, and M=6.95, SD=1.20 in S+), $t(42)$=-0.99, *ns*.

[4] A higher Representative Rating confirmed this.
[5] Comparable Representative Ratings confirmed this.
[6] This could be the effect of choosing images that were comparably arousing and inducing similar levels of feelings of dominance.

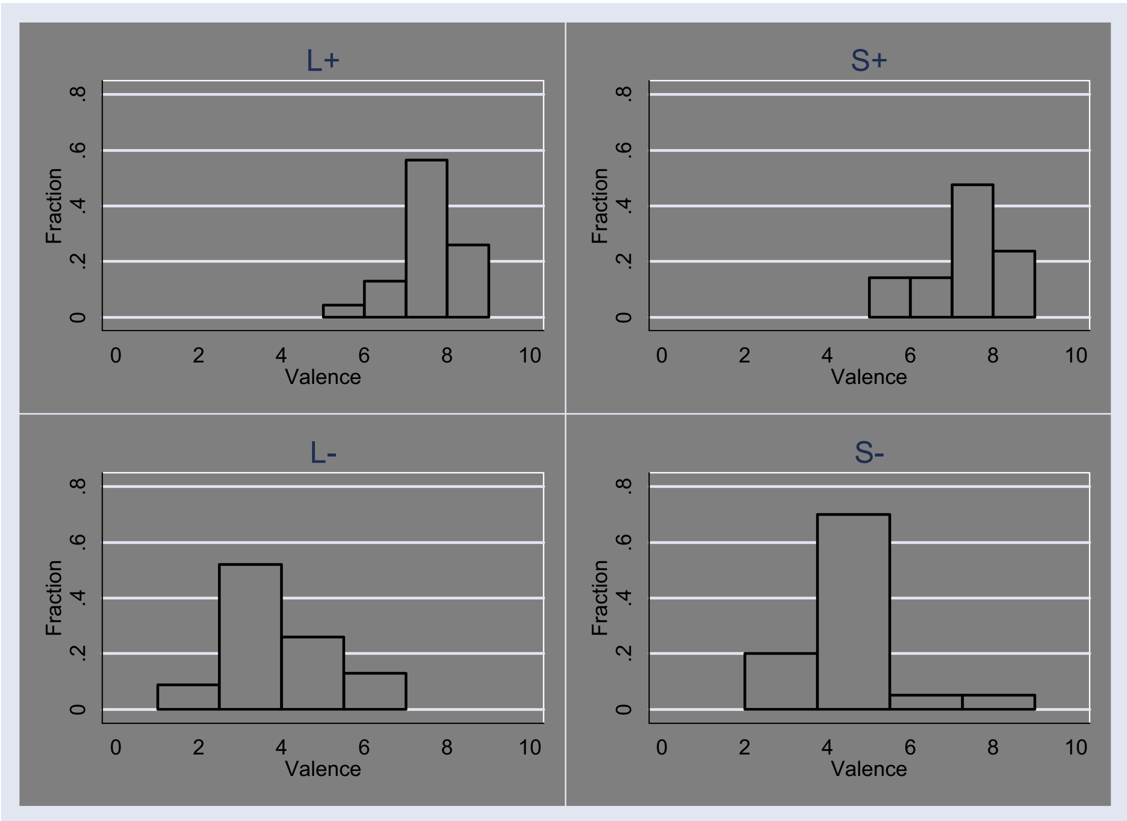


Figure 3.3: Histogram of valence ratings in experimental treatments.

Table 3.3: Zero-order correlations between overall evaluations, features of experiences and mood valence.

Features	Overall (3-item scale)	Rep. Rating	Mood Valence	Average Affect	Peak Affect	End Affect
Pleasant experiences						
Rep. Rating	.62					
Mood Valence	.61	.28				
Av. Affect	.59	.76	.21			
Peak Affect	.58	.70	.35	.70		
Duration	.30	.21	.15	-.03	.35	-.27
Variance	.27	.21	.38	.31	.51	.29
End Affect	.23	.39	-.03	.62	.32	
Slope	-.14	-.01	-.10	.10	.06	.58
Unpleasant experiences						
Rep. Rating	.50					
Mood Valence	.44	.47				
Av. Affect	.60	.51	.18			
Peak Affect	.28	.48	.39	.36		
Duration	.06	.07	-.30	.40	-.46	.24
Variance	.32	-.02	.26	.18	.05	.05
End Affect	.35	.29	.10	.53	.14	
Slope	-.03	-.18	-.05	-.06	-.18	.44

Correlations above .30 significant at 5%, and above .39 at 1%.

Thus, viewing experiences did affect mood states. L+ and L– participants were in moods that were moderately good and depressed respectively. Mood seems to have been affected to a greater degree after experiences that lasted longer, at least in the case of unpleasant experiences. However, the effect is confounded by the unintended difference in peak affect (see Table 3.2) [7]. The analysis of zero-order

[7] It was not intended that peaks in long series be more intense than peaks in short series. The image intended to evoke peak affect was present in both short and long series of images. I verified that this image evoked similar affect intensity. However, additional images in long series must have been the source of a different peak. This happened to be the case for both pleasant and unpleasant experiences

correlations between mood valence and features of experiences revealed that mood valence correlated highest with overall evaluations in both pleasant and unpleasant experiences (see Table 3.3). The overall evaluation correlated highest with the average momentary affect. And yet, it was peak, rather than average affect, that correlated highest with mood valence. While mood valence correlated positively and significantly with affect variance in pleasant experiences, in unpleasant experiences it did not.

I used multi-variable regression to examine partial effects of the features of experiences on mood and examine whether these were mediated by overall evaluations. Tables 3.4–3.5 present the results for pleasant and unpleasant experiences. These results are robust to the use of alternative measures of overall evaluations (3-item scale *vs* Representative Rating).

Pleasant experiences. Affect variance was the only variable to exert a significant partial effect on mood in pleasant experiences (column 3). Only 27% of variability in mood valence was explained by the model, in which features of the experience, but not its overall evaluation, were the independent variables. As subsequent analysis showed, the effect of affect variance was fully mediated by the overall evaluation of the experience. The overall evaluation was a significant predictor of mood valence. As a single predictor, it helped explain 37% of variability in mood valence (column 2). When introduced into the regression in addition to other features of the experience, it was the only variable to have a significant partial effect on mood. The percentage of variability in mood valence ratings explained by the model almost doubled (column 4), from 27% to 50%.

Another regression analysis, in which the overall evaluation was the dependent variable and features of the experience independent variables, showed that average momentary affect, affect variance and the duration of the experience were statistically significant predictors of overall evaluations. The model, including these and other features of the experience (peak affect, end affect and slope) explained 54% of variability in overall evaluations. Given that average momentary affect and the duration of the experience were important to explain overall evaluations, but had no indirect effects on mood, while overall evaluations themselves did predict mood, important determinants of overall evaluations seem to be missing from the equation.

Unpleasant experiences. No feature of the experience considered had a significant partial effect on mood in the case of unpleasant experiences (column 3). Average momentary affect was not the exception despite being a significant predictor of the overall evaluation (column 1). The overall evaluation, in its turn, was a significant predictor of mood valence, both by itself (column 2) and when controlling for other features of the experience (column 4). In the latter case, it was the only variable besides the constant term to have a significant partial effect on mood. Its addition to the model improved model fit from 23% to 33%. As in the case of pleasant experiences, important determinants of overall evaluation that could be contributing to mood valence were not captured among the features of experiences used in the analysis. Compared to the case of pleasant experiences, models considered for the explanation of overall evaluations and mood valence in unpleasant experiences had inferior fits.

Results of multi-variable regressions suggest that features of experiences did not explain mood valence (affect variance in pleasant experiences was an exception).

Table 3.4: Pleasant experiences: testing hypothesized effects of features of experiences on overall evaluations and mood valence.

Dependent Variable	Overall (3-item scale)	Mood Valence	Mood Valence	Mood Valence
	(1)	(2)	(3)	(4)
	Coefficient (t-statistic)	Coefficient (t-statistic)	Coefficient (t-statistic)	Coefficient (t-statistic)
Predictors				
Overall (3-item scale)		1.303		1.485
		(5.01)**		(4.00)**
Av. Affect	.019		.021	-.008
	(2.77)**		(1.12)	(0.45)
Peak Affect	-.006		-.013	-.004
	(-.90)		(.71)	(.24)
End Affect	.001		-.005	-.007
	(-.35)		(.72)	(1.08)
Slope	-.006		-.004	.004
	(-1.40)		(.39)	(.44)
Variance	.000		.001	.001
	(2.13)*		(2.53)*	(1.51)
Duration	.545		.833	.024
	(2.74)**		(1.57)	(.05)
Constant	-.143	6.047	6.035	6.248
	(-0.55)	(23.30)**	(8.68)**	(10.60)**
Observations	44	44	44	44
R-squared	.54	.37	.27	.50
Adj. R-squared	.46	.36	.15	.40

* significant at 5%; ** significant at 1%

Table 3.5: Unpleasant experiences: testing hypothesized effects of features of experiences on overall evaluations and mood valence.

Dependent Variable	Overall (3-item scale)	Mood Valence	Mood Valence	Mood Valence
	(1)	(2)	(3)	(4)
	Coefficient (t-statistic)	Coefficient (t-statistic)	Coefficient (t-statistic)	Coefficient (t-statistic)
Predictors				
Overall (3-item scale)		1.774		1.672
		(3.17)**		(2.29)*
Av. Affect	.013		.008	-.013
	(2.42)*		(.33)	(.53)
Peak Affect	-.003		.023	.028
	(0.42)		(.76)	(.96)
End Affect	.002		.005	.003
	(.84)		(.57)	(.28)
Slope	-.003		-.011	-.006
	(.83)		(.66)	(.38)
Variance	.000		.000	.000
	(.99)		(.92)	(.59)
Duration	-.174		-.707	-.415
	(.86)		(.75)	(.46)
Constant	-.573	5.775	6.678	7.636
	(1.60)	(10.98)**	(4.03)**	(4.70)**
Observations	43		43	43
R-squared	.43	.20	.23	.33
Adj. R-squared	.33	.18	.11	.20

* significant at 5%; ** significant at 1%

Overall evaluations of experiences did. Features of experiences explained overall evaluations imperfectly, and more has to be learnt about determinants of overall evaluations that were missing from the present analysis. These appear to have been important for mood induction.

3.5.3 Mood effects

Several comparisons between treatments L+ and L– are illustrative about the opening argument of this work concerning the importance of moods. Mood induction in the present experiment has led to the following effects. In the spirit of the phenomenon of mood-congruent recall, the contents of the most pleasant/least unpleasant image were recalled by a significantly greater number of participants in treatment L+ compared to treatment L– (20 of 23 in L+ versus 12 of 23 in L–), z=2.58, $p < 0.01$. This did not occur in the case of recall for contents of the least pleasant/most unpleasant image (17 of 23 in L+ and 13 of 23 in L–), z=1.23, *ns*. Note, that recall of the contents for both the most pleasant/least unpleasant and least pleasant/most unpleasant image was better in treatment L+. Meanwhile, recall in treatment S+ was comparable to that in S–. Recall of the contents of the first and the last image in treatments L+ and L– did not differ significantly.

When asked to report perceived average of their image ratings, participants were most accurate in treatment L– (depressed mood), i.e. mean absolute deviation from true average was significantly lower in L– (M=7.35, SD=5.53 in L– and M=12.48, SD=14.40 in L+), $t(44)$=1.60, $p < 0.10$. In treatments S+ and S– (only 3 images viewed), performance was similar to that in L–. These results concerning differences between treatments L+ and L– are robust to comparing only the clearly happy in L+ (mood valence of 7 and above, $N = 19$) to the clearly depressed in L– (mood valence of 3 and below, $N = 14$).

3.6 Discussion

Research has already tackled the problem of designing experiences that would be remembered as more satisfying. The big question that had motivated the present work has been whether better movies, more pleasant commercials and less painful medical procedures impact moods of customers, spectators and patients besides improving experience-specific overall evaluations.

I found that mood valence depends positively on overall evaluation of experiences. However, the features of momentary affect that explain an important portion of variability in overall evaluations did not affect moods. In pleasant experiences, only variance in momentary affect had a significant indirect effect on post-experience mood, and none of the features considered did so in unpleasant experiences. Unknown factors, those that would explain the remaining portion of variability in overall evaluations, seemed to have been responsible for the intensity of post-experience mood.

Of course, present findings are conditional on the magnitude of variation in the features of experiences in the experiment, and this can be inspected in Tables 3.1–3.2. Pre-experience mood was assumed to be neutral, however, ideally, it should have been measured. Evidence from web-based pilot studies I conducted in the past shows that pre-experience moods are often concentrated around positive ratings. In two studies, in which participants had to express their agreement with the statement "My mood is ideal right now" using a 7-point Likert scale, 11 of 17 participants in one (65%) and 12 of 14 in the other (86%) gave ratings between 5 and 6.

Moreover, the assumption in this work has been the ability of an individual to distinguish between his/her perceived momentary affect, the overall evaluation of the experience, and a referent-free mood state. Additionally, it was assumed that momentary affect was stimulus-specific, and overall evaluations and moods arose/became accessible to consciousness only post-experience. These are important assumptions that may require revision in future research. Mood dimensions of arousal and control as a function of stimuli-specific arousal and control could be studied in future research as well.

Finally, while the interest in the present work concerned the effects of experiences on moods, mood, in its turn, could affect experiences. Although not necessarily so (Forgas, 1995; Wegener et al., 1995), overall evaluations of experiences may be subject to mood effects. Schwarz and Clore (2003) advanced a Feelings-as-Information hypothesis, arguing that moods inform the decision-maker about the environment. Environments that feel "bad" are likely to be dangerous and require greater attention and a more thorough processing of information. In numerous experimental settings individuals in a sad mood were more likely to use a systematic, data-driven strategy of information processing, with considerable attention to detail. In contrast, individuals in a happy mood were more likely to rely on preexisting general knowledge structures, pursuing a top-down, heuristic strategy of information processing, with less attention to detail. In impression formation tasks, individuals in a sad mood were shown to make more use of detailed individuating information and found to be less influenced by the order of information presentation (Schwarz & Clore, 2006). Another important mood-related phenomenon is the better recall of mood-congruent information from memory. In the context of performance appraisals, for example, individuals in good moods were found to recall more positive information about performance than were individuals experiencing more neutral or negative moods (Robbins & DeNisi, 1998). Although mood effects were studied for experiences of mixed affective valence, research has not yet tackled the implications of moods for aggregating affect intensity in experiences that are pleasant or unpleasant throughout.

Chapter 4

Lay Intuitions about Overall Evaluations of Experiences

4.1 Introduction

People often report experiences by expressing a number on a scale. Someone might say, "7 out of 10 for this concert", or "In terms of painfulness, I rate this medical procedure as 90 out of 100." Such overall evaluations of experiences have been shown to be important decision inputs (Wirtz, Kruger, Napa Scollon, & Diener, 2003; Oishi & Sullivan, 2006), and studied extensively.

Kahneman et al. (1997) suggested that experiences can be represented as intensity profiles of pleasure (or discomfort) over bounded intervals of time, i.e., time profiles of "experienced utility". They cited Edgeworth who was hoping that special devices called "hedometers" would be able to measure experienced utility. Experiments and field studies have shown that people evaluate more positively experiences with increasing, rather than decreasing time profiles at equivalent levels of total pleasure experienced (Loewenstein, 1987; Ariely & Carmon, 2000). There is a preference for steeper rates of improvement (Hsee & Abelson, 1991), as well as variability in experience (Read et al., 1999). Finally, the "Peak-End rule" finding suggests that overall evaluations are best predicted by only two moments of the experience: the most pleasant/unpleasant and final (Kahneman, 2000). Kahneman et al. (1997) present a set of assumptions about experiences explaining why integration/summation of all moments would be correct from a normative point of view. Life satisfaction researchers and psychologists, on the other hand, explore alternative paradigms and study the role of personality and the beliefs of the evaluating person for overall evaluations (Updegraff, Gable, & Taylor, 2004; Robinson & Clore, 2002; Trope & Liberman, 2003; Brendl & Higgins, 1995).

In contrast to previous research, the present work aims to reveal what decision makers themselves draw on as they think about overall evaluations of experiences. I will compare lay intuitions to what researchers have considered. This comparison may further enrich theories of overall evaluations, and suggest ways of testing them.

I employ a novel method, the guessing task, in order to elicit lay intuitions. The philosophy of the method is that of Active Information Search, a method of naturalistic decision-making that Huber, Wider, and Huber (1997) proposed for the study of risky choice. The method consists of giving participants a minimal description of a decision problem and allowing them to seek information. I report experiments in which participants had to guess the overall evaluation of an experience lived by another person. Active information search was allowed prior to making the guess. The information participants sought was taken to reveal lay intuitions about the target overall evaluation.

4.2 General method

4.2.1 The guessing task

Participants were assigned randomly to be *Informants* or *Guessers*.

Informant's task. Informants lived and evaluated a certain experience. Their evaluations were unknown to Guessers. These were ratings on 0 to 100 point scale anchored by statements about experienced pleasure or discomfort. Ratings were real-time and overall. Real-time ratings ranged from 0, "Not pleasant at all", to 100, "Very pleasant", and overall from 0, "I experienced no pleasure at all" to 100,"I experienced a great deal of pleasure". For example, if an Informant listened to several musical performances, he/she evaluated each performance immediately after hearing it and the musical sequence overall. If an Informant tasted pieces of chocolate, he/she rated each piece and then rated the whole tasting session. If an Informant viewed affective images, he/she rated each image and then the experience of viewing the whole series. [1] Informants wrote on evaluation sheets distributed to them prior to the experience.

Guesser's task. A Guesser was a participant who faced the task of guessing the overall rating that an Informant gave to his/her experience. Guessers could not communicate with each other. They knew the class of stimuli experienced by the Informant, but not the duration of the experience (e.g., that the Informant had listened

[1] In the experiment involving the viewing of aversive images, the anchoring statements for real-time and overall ratings referred to discomfort instead of pleasure

to musical performances, had tasted chocolate samples, etc.); and they knew that the Informant rated the experience in real-time and overall using 0 to 100 point scales with anchoring statements "Not pleasant at all"/"I experienced no pleasure at all" and "Very pleasant"/"I experienced a great deal of pleasure".

Guessers could ask Informants questions. They were instructed to refrain from judging the appropriateness of questions. [2] Questions had to be written down and could be asked simultaneously or sequentially.

4.2.2 Closed-format questionnaires

Closed-format questionnaires complemented the guessing task. Questionnaire A was designed prior to Experiment 1, and contained questions about the experience of the Informant inspired by the "hedometer" perspective; that is, the items were questions about real-time ratings and statistics of these. Questionnaire B was designed after Experiment 1 for participants of subsequent experiments. The items were exemplars of question categories observed in Experiment 1. Each Guesser faced a different order of items in each questionnaire. Guessers were asked to pick three questions among questionnaire items that would be most useful in the guessing task, and underline the most informative one.

[2] Direct questions of the type "What was your overall evaluation?" or "Was your overall rating below 50?" were not transmitted to the Informants, and Guessers were asked to formulate a different question. If attempted repeatedly, such questions were allowed finally, but a special classification category was created for them. Questions of this category did not exceed 9% of total questions in any of the experiments/conditions.

4.3 Experiment 1

4.3.1 Method

Stimuli

Informants were exposed to either a short or a long auditory experience, consisting, respectively, of two or six Moldovan folk music performances. Each performance was 2-4 minutes long. Informants had earphones. Windows Media player was used to reproduce the music.

Participants

There were 54 participants in Experiment 1; 22 were male, average age was 22. All were undergraduates, and the vast majority were students in economics. 18 participants acted as Guessers, of which 6 were male, average age was 21.

Procedure

Informants were assigned randomly to 2 versus 4 performances. The number of questions that a Guesser could ask was limited to 3 when the Informant had listened to 2 performances, and to 5, when the Informant had listened to 4 performances.

Informants and Guessers were paid 2 euros for participating. For each guess where the error was less than 5 points the Guesser received 5 euros.

There were 18 sessions in total, each involving a Guesser and two Informants (one Informant was involved in a separate task, and his role is not discussed in this article). In each session, the Informant had evaluated his/her experience prior to the arrival of the Guesser. The Guesser wrote down his/her questions. The author passed the questions of the Guesser to the Informant and delivered answers back to the Guesser. Once the Informant had answered all questions, the Guesser made his/her guess. Guessers wrote down a comment explaining the guessing strategy used in a short post-guess questionnaire. They wrote down additionally the question they would have asked had they been constrained to a single question, and completed questionnaire A. Finally, they received performance feedback.

Classification and coding of questions

I content analyzed the questions in view of the perspectives on experiences that the questions reflected. For example, questions involving real-time ratings were consistent with the "hedometer" perspective on experiences discussed earlier. Other perspectives involved the Informant's personality, the category of the experience, other holistic attributes and judgments, as well as perceived behavioral implications. Classification codes for question types within each perspective, and examples, were formulated for use by independent coders (there were 14 codes). Finally, one female graduate student in clinical psychology and one male graduate student in economics coded the questions. There were no disagreements. In what follows, I present the resulting classification of questions, discussing the codes within 5 broad categories: "hedometer" perspective, holistic attributes and judgments [holistic A/J], Informant's personality, decision rule, and behavioral implications.

Category 1: "hedometer" perspective. The first category comprised all questions inquiring for real-time ratings and any statistics of these [ratings stats].

These were questions of the type: "How did you rate the musical performances of this sequence?" or "How did you rate the performance you liked the best?". Importantly, there was not a prevalence of questions about the maximum/minimum or final ratings. Guessers also asked for the average and modal ratings, the trend, the slope, and the variance of ratings.

Category 2: holistic attributes and judgments [holistic A/J]. The second category included questions such as, "What was, or, how much did you like the rhythm of the music you listened to?", "Was the music you heard classical?", "Was your experience with music similar to your experience in a philosophy lecture?", or "Did you feel tender emotion as you listened to this music?". As the latter questions suggest, specific emotions or the category of the experience indicated the experience's overall evaluation. Notable are questions that referred to holistic attributes of the experience identifiable only in retrospect (i.e., the overall rhythm of the music).

Category 3: personality. The third category included questions related to the personality of the Informant. Guessers asked about social status, general knowledge and culture, as well as enduring psychological dispositions. For example, "Are you a person who likes variety?", "Are you a generally depressed individual?".

Category 4: decision rule. The fourth category involved inquiries about the decision rule underlying the overall utility rating, for example, "Did you rate the experience overall based on the fact that you are generally fond of music or rather based on your actual experience with these pieces of music (that is your overall rating was equal to the average of piece ratings)?".

Category 5: behavioral implications. The fifth category comprised questions, which explored the implications of a given overall rating for the experience's future use, willingness-to-pay (WTP) for it, what purpose the experience could serve and how useful it would be. For example, guessers asked "Could you use this music as a background for a romantic dinner?", or "How often would you listen to this music if you had it at home?".

Table 4.1 reports both the proportion of a question category in total questions asked (TQ, an idea's "persistence") and the proportion of participants asking questions of a particular category (AP, an idea's "spread"). A given Guesser often asked questions pertaining to different categories. Therefore, I calculated proportions of participants asking a particular combination of question categories, and report these additionally. Table 4.2 reports the structure of single questions that Guessers wrote down.

4.3.2 Results and discussion

Guessers formulated a total of 66 questions. Note, that the instructions could direct their thinking towards a "hedometer" perspective on experiences (they were told explicitly that Informants had rated their experiences in real-time and overall, and could think that the two types of ratings had to be related). Importantly, research adopting a "hedometer" perspective provides an indication of how to use real-time ratings for predicting overall evaluations. One strategy is to compute the average of the most extreme and final real-time rating, and another - to compute the over-all average (Kahneman, 2000; Ariely & Zauberman, 2003; Langer et al., 2005). The results of the guessing task show that the first strategy was not intuited by Guessers. The

Table 4.1: Content structure of guessers' multiple questions: proportion of total questions (TQ, %), and proportion of participants asking the question of a particular type (AP, %).

Category and codes	Exp. 1 (music)		Exp. 2 (chocolate)		Exp. 3 (pleasant images)		Exp. 4 (aversive images)	
	TQ	AP	TQ	AP	TQ	AP	TQ	AP
"Hedometer"	**39**	**61**	**18**	**44**	**15**	**33**	**16**	**27**
– Ratings stats	20	44	7	13	8	21	16	27
– Real-time ratings	14	28	4	13	0	0	0	0
– Duration	5	17	7	22	7	17	0	0
Holistic A/J	**29**	**61**	**38**	**74**	**39**	**67**	**47**	**100**
– Attributes	5	11	7	13	15	33	17	42
– Attribute evalns	3	11	1	4	1	4	3	8
– Experience categ.	11	39	4	13	6	13	9	27
– Category liking	5	11	25	61	3	8	0	0
– Emotions	5	17	1	4	14	38	18	46
Beh. Impl.	**19**	**39**	**20**	**44**	**26**	**63**	**20**	**54**
– Future use	8	17	15	30	10	29	8	23
– Instrumentality	8	17	4	13	13	29	7	15
– Approach/avoid.	0	0	1	4	3	8	4	12
– WTP	3	11	0	0	0	4	1	4
Personality	**14**	**22**	**6**	**17**	**10**	**25**	**7**	**19**
Decision rule	**2**	**6**	**6**	**9**	**1**	**4**	**1**	**4**
Non-classif.	*2*	–	*9*	–	*8*	–	*9*	–
Total /All	N=66	N=18	N=68	N=23	N=72	N=24	N=76	N=26

Table 4.2: Content structure of guessers' single questions: proportion of total questions (%).

Category and codes	Exp. 1 (music)	Exp. 2 (chocolate)	Exp. 3 (pleasant images)	Exp. 4 (aversive images)
"Hedometer"	**39**	**13**	**30**	**15**
– Ratings stats	22	9	17	15
– Real-time ratings	17	4	0	0
– Duration	0	0	13	0
Beh. Impl.	**39**	**52**	**29**	**20**
– Future use	33	35	8	8
– Instrumentality	0	9	21	4
– Approach/Avoid.	0	4	0	4
– WTP	6	4	17	4
Holistic A/J	**12**	**34**	**34**	**58**
– Attributes	0	0	17	23
– Experience categ.	6	17	4	12
– Category liking	6	17	0	0
– Emotions	0	0	13	23
Personality	**6**	**0**	**0**	**0**
Decision rule	**0**	**0**	**4**	**0**
Non-classif.	*6*	*0*	*4*	*8*
Total questions	N=18	N=23	N=24	N=26

second was pursued in some cases. It was preferred to other strategies involving real-time ratings, as closed-format questionnaires also showed (see Table 4.3). Duration of the experience was rarely a matter of concern to the Guessers.

Although the largest proportion of questions in total questions asked revealed a “hedometer” perspective, questions involving holistic A/J were equally important in spread, i.e. in terms of the proportion of participants asking at least one question of the kind. In addition, Guessers asked frequently about behavioral implications of the overall rating and the personality of the Informant. Table 4.2 describes the structure of single questions formulated by Guessers, and shows that most question categories remained represented in a similar order of importance.

The analysis of how participants combined frames of analysis reveals that most of them asked questions pertaining to at least two question categories (66%). Most frequent types of combinations involved the “hedometer” perspective and holistic A/J, or the latter and behavioral implications. 22% maintained a “hedometer” perspective.

Questions Guessers asked helped them make 10 successful guesses in 18 attempts. If they had pursued the strategy of averaging across particular, or all real-time ratings, as described above, the success rate would have been 13 in 18.

4.4 Experiments 2, 3 and 4

One could argue that a musical experience is different from other hedonic experiences, such as food tasting or pain. I report replications involving tasting chocolate, and image-viewing experiments, which allowed experimentation with both pleasant and aversive stimuli.

Table 4.3: Questionnaire A. Proportion of guessers choosing an item (%).

Questionnaire items*	Exp.1 (music)	Exp. 2 (chocolate)	Exp. 3 (pleasant images)	Exp. 4 (aversive images)
Average rating	67	65	63	65
Maximum rating	56	35	42	15
Minimum rating	56	9	21	15
Modal rating	39	52	63	73
Sum of ratings	17	26	42	38
Trend	na	39	17	31
Variance	na	35	25	31
Middle rating	11	4	0	4
End rating	11	13	4	8
Duration	11	17	42	23
Location of maximum	11	4	0	15
First rating	6	4	0	0
Total participants	N=18	N=23	N=24	N=26

*Questionnaire items were questions formulated so as to avoid the use of statistical terms. For example, the question about the modal rating read: "What was the most frequent rating you used to rate these performances/pieces/images?", the question about the trend: "Was the experience increasingly pleasant or increasingly unpleasant?", the question about the maximum rating: "What was the rating of your favorite performance/piece/image?", and so on. Preferences in favor of items chosen by at least 35% of participants are significant at 10% and higher levels of statistical significance.

4.4.1 Method

Stimuli

In Experiment 2 Informants tasted two or six pieces of chocolate. Pieces were small portions of white, black, milk, baking, liquor filling, and nuts and raisins chocolate. Each piece was presented to the taster on a napkin and covered by a napkin with the number indicating its order of tasting and no other information.

Two and 15 pleasant, and two and 15 aversive images were the stimuli for viewing experiences in Experiments 3 and 4 respectively (Lang et al., 2005). Images appeared for 7 seconds each in a PowerPoint presentation.

Participants

There were 27 participants in Experiment 2, of which 23 acted as Guessers (11 were male, average age was 21). There were 28 participants in Experiment 3, of which 24 acted as Guessers (12 were male, average age was 21), and 30 participants in Experiment 4, of which 26 acted as Guessers (9 were male, average age was 20). Participants were undergraduate students in diverse disciplines. [3]

Procedure

There were two conditions: in condition 1 Guessers could ask three questions; and in condition 2 only one. Guessers were assigned randomly to two orders of conditions. In each condition they addressed a different Informant and had one guessing attempt. Guessers were rewarded for accurate guesses in one condition chosen at random (an

[3] Law, political science, economics, humanities, biology, engineering, computer science, journalism, management, and several related disciplines.

error of 7 points was allowed) . If the guess had been made based on three questions, Guessers were paid 10 euros. If it had been made based on one question they were paid 15 euros. The show-up fee was 12 euros for Informants and 5 for Guessers. [4]

There were two new items added to questionnaire A: questions about the trend and the variance of real-time ratings. Questionnaire B contained 9 items representative of question categories observed in Experiment 1 (see Table 4.4).

Every experiment was conducted in two 50-minutes sessions. Prior to the beginning of each session, 2 participants had to prepare for the role of Informants. One experienced the long version of the experience, and the other the short version. Both Informants had rated experiences lived and were ready to reply to the questions of Guessers by the time the session began. Experiments were run in spacious classrooms with Informants seated in the back rows and at a distance from Guessers. Performance feedback was given after Guessers completed both conditions and answered questionnaires A and B.

4.4.2 Results and discussion

Tables 4.1 and 4.2 report the structure of questions in Experiments 2-4. [5] It is consistent with previous findings. [6]

The analysis of how participants combined frames of analysis revealed that most of them combined at least two (65% of all participants in Experiment 2 (chocolate

[4] A new recruitment system established higher minimum hourly pay rates.

[5] Classification scheme developed in Experiment 1 was used to categorize the questions of Guessers. The inter-coder agreement was 94% in Experiment 2, 85% in Experiment 3, and 86% in Experiment 4.

[6] A proper statistical qualification of the differences would require a greater sample, perhaps, a different method, and lies outside the scope of the present article.

tasting), 58% in Experiment 3 (positive images) and 81% in Experiment 4 (aversive images)). Most frequent types of combinations in Experiment 2 were equivalent to those observed in Experiment 1. Most frequent types of combinations in Experiments 3-4 involved holistic A/J and behavioral implications. In Experiment 4 two other important combinations involved the "hedometer" perspective and holistic A/J, as well as the "hedometer" perspective and personality. The most frequent single frame of analysis involved holistic A/J.

In Experiment 2, there were 26 successful guesses in 48 attempts in two conditions. If participants had asked for real-time ratings only, and averaged them, or computed the average of maximum and final ratings, their performance would not have improved (24, or 12 guesses would have been made). In Experiments 3-4, instead of 15 successful guesses in 48 attempts and 13 in 52, participants would have attained higher success rates of 37 in 48 and 42 in 52 by averaging across real-time ratings, or, been correct half of the time if believing in the "Peak-End rule".

Choices of items in questionnaire A resembled results obtained in Experiment 1. Guessers believed that average and modal real-time ratings were the most informative pieces of information about the experience's overall evaluation. Choices of items in Questionnaire B showed that the importance of average real-time ratings withstood the comparison to other kinds of questions. In Experiments 2-3, Guessers thought that the willingness-to-pay for the experience and the knowledge of experience category is comparable in importance to the knowledge of average real-time rating.

Table 4.4: Questionnaire B. Proportion of guessers choosing an item (%).

Questionnaire items	Exp. 2 (chocolate)	Exp. 3 (pleasant images)	Exp. 4 (aversive images)
– What was the average rating you had given to these pieces/images?	65	67	77
– How much would you pay to repeat this experience?	52	25	19
– What was the rating of your favorite piece/image?	48	54	69
– How many times per month would you like to repeat this experience?	48	25	8
– What other experience would you describe as similar?	43	46	77
– How would you rate your life in general?	13	38	19
– Would this experience be useful to entertain friends?	13	17	15
– How many pieces/images did you sample/view?	9	33	15
– How would you rate the sweetness/color scheme of the samples/images?	4	8	12
Total participants	N=23	N=24	N=26

*Preferences in favor of items chosen by at least 43% of participants are significant at 10% and higher levels of statistical significance.

4.5 General discussion

Overall evaluations of experiences have been studied extensively by economists, psychologists, and philosophers. Opposing the "hedometer" perspective on experiences, the latter have argued for the role of "reconstruction" in overall evaluations (Alexandrova, 2005). By means of a novel guessing task with Active Information Search, I identified the considerations that decision-makers themselves relate to overall evaluations. These may play an important role in the process of "reconstruction", as well as in interpersonal communication on the subject.

Researchers may be interested to learn whether certain features of overall evaluations are intuited by people. For example, most theories and empirical findings suggest that the duration of experiences is not an important determinant of overall ratings. Importantly, lay theorists manifested similar beliefs by paying little attention to duration in the search of information about overall evaluations. However, while researchers distinguish between overall ratings and willingness-to-pay judgments (Ariely & Loewenstein, 2000), some lay theorists confounded the two.

Frames of analysis employed by lay theorists have been shown to parallel frames of analysis in academic theorizing on subjective satisfaction judgments used within separate research traditions. However, multiple frames were evoked simultaneously in the minds of lay theorists with respect to very simple experiences. This suggests the need to explore potential interactions. Moreover, different people have used different frames. Thus, future research can be aimed at exploring features of the communication context that allow people to coordinate on a given frame and a meaning for overall evaluations of experiences.

Methodologically, this work contributes to the study of human judgment by demonstrating what lay intuitions can add to laboratory findings and the assumptions of a research tradition. The guessing task can be used for the study of lay intuitions in a number of settings, from the forecasting of preferences for specific objects to the predictions of actions in situations of strategic interaction. Importantly, Active Information Search can then be easily made incentive-compatible, and allow the manipulation of stakes involved. The separation between the target of the prediction and the Guesser provides additionally the possibility of exploring self-other differences in evaluation criteria, beliefs and the framing of many decision situations. Even when participants are not able to articulate their intuitions perfectly, the researcher is able to document general frames of analysis employed and, therefore, the fundamentals likely to be used in further articulating a lay theory.

Chapter 5

Why Not Learn the Peak-End Rule?

5.1 Introduction

Imagine, the authors of this chapter [1] went to a concert of classical music. Later, each told you how much he/she enjoyed every musical piece heard. This was done using a scale from 0 for "didn't like it at all" to 100 for "liked it very much". Given such information, would you be able to tell how piece evaluations determine the overall evaluation of the concert for each of us? Perhaps, you have a cd of the concert, and wonder whether one of us would be happier to receive it as a gift.

A concert, a day at work, an interview – all are examples of experiences lived across time. Previous research has shown that when people *aggregate* their experience, i.e. provide a single overall evaluation of a series of impressions, they only account for the

[1] This work was done in collaboration with Dmitry Ryvkin, Department of Economics, Florida State University.

most extreme and final impressions, Peaks and Ends (Kahneman, 2000). The rest of the experience is virtually neglected.

The so-called "Peak-End rule"(the average of only Peak and End impressions) has been shown to underlie the overall evaluations of hypothetical pain episodes, lives, non-hypothetical experiences in the laboratory involving annoying sounds, pain from the immersion of hands into cold water, painful pressure from a vice, the viewing of pleasant and aversive plotless video-clips, advertisements, episodes of social interaction, receipt of money payments and music experiences, and experiences in fields such as hospitals and queues (Varey & Kahneman, 1992; Diener et al., 2001; Schreiber & Kahneman, 2000; Ariely & Loewenstein, 2000; Kahneman et al., 1993; Ariely, 1998; Fredrickson & Kahneman, 1993; Baumgartner et al., 1997; Fredrickson, 2000; Langer et al., 2005; Rozin et al., 2004; Redelmeier & Kahneman, 1996; Ariely & Carmon, 2000; Carmon & Kahneman, 1996). So, for example, the person who enjoyed the concert the most is the one who rated his/her favorite piece and the piece that closed the concert the highest.

The Peak-End rule implies that overall evaluations of experiences do not account for the duration of experiences and virtually neglect many relevant impressions. It is for these reasons, that Kahneman et al. (1997) have qualified the Peak-End rule as dysfunctional. They argued that overall evaluations should reflect the sum or the temporal integral of impressions experienced. This argument was supported by a set of axioms and conditions hypothesized to underlie impressions of satisfaction in experiences lived across time.

In a series of experiments, Cojuharenco (2007) explored folk beliefs about overall evaluations of experiences. One of the goals was to find out whether participants were aware of the Peak-End rule, and understood/anticipated the loss of information involved in overall evaluations of experiences. Participants believed, it was found, that overall evaluations reflected the average of *all* impressions. So, they would be tempted to infer our overall evaluations of the concert by averaging across piece evaluations, rather than considering the favorite and the final pieces only.

The finding was consistent with a number of previous indications in the literature. Ariely (1998) noted that experimental participants tended to average their impressions when evaluating experiences whenever these impressions were documented online, and, therefore, readily available to them. Baumgartner et al. (1997) pointed out that advertisement researchers assumed commonly that the overall ad liking reflected the average momentary reaction to the ad. Furthermore, evaluations of equal-length experiences by average impressions would be consistent with what has been considered normatively correct, i.e. the temporal integral of impressions (Kahneman et al., 1997).

Overall evaluations of experiences and tasks of the kind we suggest in the opening example abound in daily appraisals. And yet, people have inaccurate beliefs about overall evaluations. Why not learn the Peak-End rule?

We answer this question by analyzing the structure of information in the learning environment. The Peak-End rule ranks experiences in a certain way. This ranking correlates strongly with the ranking of experiences based on their overall evaluations, i.e. the Peak-End index of the experience is a good predictor of its overall evaluation. On the other hand, the average of impressions experienced leads to an alternative ranking. By computing correlation coefficients between the Peak-End and average impressions, we operationalize the similarity between the two evaluation rules. This similarity represents a measure of information redundancy in the learning environment. Both Peak-End and average impressions could predict overall evaluations, and the greater the redundancy - the more likely so. But if both could, it would be more difficult to perceive differences in the predictive ability of each of these rules. In a meta-analysis of learning studies, Karelaia and Hogarth (2007) note that redundancy is an important obstacle to learning the predictive ability of various pieces of information with respect to a given dependent variable. In this paper, we analyze the redundancy and reasons for redundancy in the information about overall evaluations of experiences.

The paper is organized as follows. Section *Information redundancy* examines the correlation between the Peak-End and average impressions based on 54 experimental and field data sets on experiences lived across time. The criteria for data collection are described in detail. We show that the correlation and, therefore, the redundancy of information about overall evaluations, is high. We point out that the high correlation found was unexpected.

The process of experiencing is responsible for the magnitude of the correlation between Peak-End and average impressions. Section *The model* introduces the theoretical framework to guide our thinking about this process. We start by considering experiences with time-separable impressions, and identical individuals. This model is shown to be inconsistent with the correlation observed. We next relax time-separability and the assumption of identical individuals by examining a lagged dependent variable model, representing processes of *anchoring-and-adjustment* and *adaptation* in experiences lived across time (Hogarth & Einhorn, 1992; Frederick & Loewenstein, 1999). The class of processes considered is shown to allow for high correlation.

In Section *Empirical analysis*, we fit the model to the available panel data, controlling for the unobserved individual-specific effects, which may further contribute to the redundancy of information. We summarize the results of this exercise. The persistence in impressions experienced and individual heterogeneity help explain why decision-makers find it difficult to distinguish what, the Peak-End or the average impressions, predict overall evaluations better.

Consider our opening example. If the evaluation of each musical piece is anchored to the piece heard previously, piece evaluations would be more similar to each other than they would be otherwise. Given this similarity, one could use any piece evaluation as the approximation of the average piece evaluation. In addition, if the authors of this paper differed greatly in their general attitudes towards classical music, the evaluation of each piece heard could be affected. Piece evaluations in line with a generally positive attitude would be high, while those in line with a generally negative attitude low. As a result, it would not matter much whether differences in overall evaluations would be summarized by using few rather than all high versus low piece evaluations.

We summarize the implications of our findings in Section *Discussion*. Our contribution is threefold. We identify reasons why learning the Peak-End rule may be difficult. We test whether known psychological processes can account for the high correlation between Peak-End and average impressions in the data on experiences lived across time. Practically, our work is informative about the conditions under which the choice of a rule for overall evaluations may not be crucial.

5.2 Information redundancy

Experimental and field studies of the Peak-End rule were the target of our data collection effort because (i) either the Peak-End rule or average impressions have been shown to predict overall evaluations of experiences in these studies, and (ii) researchers measured and documented impressions throughout experiences. Experimental studies that controlled for average impressions were not suitable for our purposes. We contacted the authors of studies that could be used to compute the correlation between Peak-End and average impressions (Baumgartner et al., 1997; Ariely & Carmon, 2003; Fredrickson & Kahneman, 1993; Redelmeier & Kahneman, 1996). Baumgartner et al. (1997) and Ariely and Carmon (2003) have kindly responded to our data request, providing a total of 35 data sets. Additional 19 data sets came from the unpublished research of Irina Cojuharenco, who has used experimental and field data to test the Peak-End rule. Finally, experiences documented ranged from the viewing of advertisements (per second evaluations) to classroom experiences (evaluations of explanations or discussions every 5-10 min), image-viewing (evaluations of images every 7 sec), pain (hourly reports of pain intensity) and evaluations of various life aspects during a month-long study of life satisfaction (reports every 3 days). The description of the data sets and the correlations found between Peak-End rule and the average of all impressions experienced are reported in Table 5.1, columns 1-5. The

Peak-End impressions of many experiences were correlated highly and significantly with the average impressions.[2]

We emphasize that we did not intentionally target or restrict our attention to the data sets that exhibit high correlation between the Peak-End and average impressions. Nevertheless, the high correlation is present in *all* data sets that came into our possession. This finding is surprising, and suggests that the content of Peaks and Ends is very similar to the content of average impressions, even when experiences involve many distinct impressions. In the following section we introduce the framework to guide our thinking about processes of experiencing that can be responsible for the high redundancy of information about overall evaluations.

5.3 The model

5.3.1 Terminology and notation

It is possible to represent experiences lived across time as series of instantaneous impressions of satisfactions that we will refer to as *moment utilities*, u_t (here and henceforth a generic subscript t is assumed to take on values $t = 1 \ldots, T$, where T is the length of the experience). Throughout this work, we will refer to the whole series of moment utilities $(u_1, \ldots, u_T)$ as the *utility profile* of the experience, or, simply, the experience.

Moment utilities are reactions to the changing exogenous states of nature or *stimuli*, s_t. Generally, moment utility u_t does not have to be determined by stimulus s_t

[2] Average correlation across data sets is 0.85. Controlling for sampling error, variation in population correlations is 0.008 (Hunter, Schmidt, & Jackson, 1982).

Table 5.1: Correlation between Peak-End and Average Total Utility in the Data (r_{data}).

Data Set	T, Exp. Length	N, Participants	r_{data}	95% Conf. Interval	$r_{\text{predicted}}$
1	30	26	0.87	0.74 – 0.94	0.93
2	30	27	0.97	0.94 – 0.99	0.98
3	30	27	0.93	0.84 – 0.97	0.98
4	30	27	0.93	0.85 – 0.97	0.91
5	30	27	0.84	0.68 – 0.93	0.94
6	30	27	0.91	0.81 – 0.95	0.97
7	30	27	0.95	0.90 – 0.98	0.98
8	30	27	0.91	0.80 – 0.96	0.94
9	30	27	0.92	0.84 – 0.97	0.93
10	30	27	0.61	0.30 – 0.80	0.94
11	42	23	0.93	0.92 – 0.94	0.98
12	45	27	0.91	0.81 – 0.96	0.98
13	45	27	0.90	0.79 – 0.96	0.97
14	45	26	0.95	0.89 – 0.98	0.96
15	50	27	0.94	0.88 – 0.97	0.94
16	50	27	0.95	0.90 – 0.98	0.98
17	52	24	0.92	0.82 – 0.96	0.97
18	57	22	0.98	0.96 – 0.99	0.97
19	60	26	0.96	0.92 – 0.98	0.98
20	60	27	0.87	0.72 – 0.94	0.97
21	60	26	0.97	0.92 – 0.99	0.94
22	60	27	0.92	0.82 – 0.96	0.98
23	60	27	0.86	0.71 – 0.93	0.98
24	60	27	0.96	0.91 – 0.98	0.99
25	60	27	0.89	0.77 – 0.95	0.97
26	60	27	0.94	0.87 – 0.97	0.97
27	60	27	0.95	0.89 – 0.98	0.98
28	60	27	0.93	0.84 – 0.97	0.97
29	73	25	0.88	0.75 – 0.95	0.98
30	75	27	0.66	0.38 – 0.83	0.90
31	90	26	0.94	0.87 – 0.97	0.98
32	90	27	0.87	0.74 – 0.94	0.97
33	90	27	0.88	0.76 – 0.95	0.97
34	90	26	0.93	0.85 – 0.97	0.99

Table 1 (continued).
Correlation between Peak-End and Average Total Utility in the Data (r_{data}).

Data Set	T, Exp. Length	N, Participants	r_{data}	95% Conf. Interval	$r_{\text{predicted}}$
35	30	23	0.50	0.11 – 0.76	0.66
36	30	23	0.73	0.45 – 0.88	0.76
37	3	20	0.74	0.45 – 0.89	N/A
38	3	20	0.75	0.47 – 0.90	N/A
39	18	42	0.80	0.66 – 0.89	0.77
40	7	36	0.80	0.63 – 0.89	0.81
41	7	46	0.76	0.61 – 0.86	0.81
42	4	42	0.81	0.67 – 0.89	N/A
43	6	28	0.67	0.40 – 0.84	0.80
44	11	37	0.84	0.70 – 0.91	0.90
45	10	35	0.90	0.80 – 0.95	0.86
46	10	35	0.83	0.68 – 0.91	0.85
47	10	35	0.90	0.82 – 0.95	0.86
48	10	35	0.89	0.78 – 0.94	0.81
49	10	35	0.91	0.83 – 0.96	0.90
50	10	35	0.62	0.37 – 0.79	0.79
51	10	35	0.72	0.51 – 0.85	0.80
52	10	35	0.69	0.46 – 0.83	0.78
53	10	35	0.69	0.47 – 0.83	0.76
54	10	35	0.79	0.62 – 0.89	0.77

Data sets $1-34$ are due to Baumgartner et al. (1997) and refer to per-second evaluations of advertisements. Data set 44 is due to Ariely and Carmon (2003) and refers to hourly reports of pain in a day-long hospital field study. Remaining data sets come from our unpublished research, data sets $35-38$ refer to evaluations of images in image-viewing experiments, $39-43$ to evaluations of classroom explanations and discussions in classroom field studies, and $45-54$ to evaluation of life aspects in a month-long life satisfaction study.

alone; it may also depend on the complete history of past stimuli $(s_1, \ldots, s_{t-1})$ and the beliefs regarding future stimuli (Manis, 1971; Ariely, Loewenstein, & Prelec, in press, 2003; Kahneman et al., 1997). Analyzing the data from an experiment in which participants viewed pleasant and aversive film clips, Fredrickson and Kahneman (1993) note, for example, that changes in real-time affect reported by the participants "were not caused by changes in film content", and that "people can discriminate endogenous processes of affective escalation and satiation from the exogenous effects of stimulus changes" (p.50). We discuss the underlying psychological processes below.

Moment utilities u_t are aggregated into a *total utility*, U (the overall evaluation). Different *aggregation rules* lead to different values of total utility for identical experiences. We will focus on two aggregation rules: the "Peak-End rule" and the "Average rule." Peak-End total utility, U_{PE}, will stand for the average between the most extreme and the final moment utilities in the experience, as opposed to Average total utility, U_{A}, accounting for all moment utilities without exception. We analyze the correlation between Peak-End and Average total utility as a measure of their similarity and redundancy of information associated with overall evaluations of experiences.

5.3.2 General setup

We consider a population of individuals $i = 1, \ldots, N$, each of whom is exposed to a series of stimuli $s_{i1}, \ldots, s_{iT}$, and thus lives through a temporally extended experience T periods long. We assume that all the stimuli are positive,[3] independent and identically distributed across individuals and time, i.e. all s_{it} are drawn independently

[3]Negative stimuli can be treated similarly. The Peak utility for positive stimuli is defined as the maximum utility, and it is the minimum for the negative stimuli (Ariely & Carmon, 2003).

from the same distribution with a finite variance σ_s^2. Individuals are heterogeneous in their demographics (such as age, gender), personality traits, and other characteristics, which affect their experience [see Di Tella and McCulloch (2006) for an example of individual-specific effects referred to as individual-specific "exaggeration"]. We model this heterogeneity by assuming that the utility of each moment contains an individual-specific time-constant term c_i ("fixed effect"). All c_i's are sampled independently from the same distribution with a finite variance σ_c^2. Following Kahneman (2000), we define the Peak-End total utility of the experience for individual i as

$$U_{\mathrm{PE}i} = \frac{1}{2}\left(u_{i\max} + u_{iT}\right), \tag{5.1}$$

where $u_{i\max} = \max_{1\leq t\leq T} u_{it}$ is the peak moment utility. Further, we define the Average total utility as

$$U_{\mathrm{A}i} = \frac{1}{T}\sum_{t=1}^{T} u_{it}. \tag{5.2}$$

Redundancy of information about overall evaluations of experiences is operationalized as the correlation coefficient between the Peak-End and Average total utilities, which is defined in the standard way,

$$r \equiv \mathrm{Corr}\left(U_{\mathrm{A}}, U_{\mathrm{PE}}\right) = \frac{\mathrm{Cov}\left(U_{\mathrm{A}}, U_{\mathrm{PE}}\right)}{\left[\mathrm{Var}\left(U_{\mathrm{A}}\right)\mathrm{Var}\left(U_{\mathrm{PE}}\right)\right]^{1/2}}. \tag{5.3}$$

Here, total utilities U_{PE} and U_{A} are the random variables corresponding to the sample quantities defined by Eqs. (5.1) and (5.2), respectively.[4] In Eq. (5.3), and throughout the analysis below, all expectation values are assumed to be taken over both the population and the stimuli. Thus, the correlation coefficient r we discuss

[4]We drop subscript i to denote random variables corresponding to sample quantities.

measures the strength of linear relationship between U_{PE} and U_{A} *in general*, not for a particular type of individuals, or a particular series of stimuli.[5]

5.3.3 Experiences with time-separable moment utilities and identical individuals

We start with the simplest case, in which the following assumptions hold.

Assumption T1: $u_t = g(s_t)$;

Assumption T2: $c_i = 0$, $i = 1, \ldots, N$.

Assumption T1 states that the moment utility at time t is fully determined by the stimulus at time t; this makes utility time-separable. Assumption T2 states that there are no differences across individuals, therefore without loss of generality we set all individual-specific effects to zero.

For illustrative calculations and simulations, we take $g(s) = s$.[6] It can be shown (for derivation see Appendix C.1) that under Assumptions T1–T2, with $g(s) = s$, and uniform distribution of stimuli, the correlation coefficient between U_{PE} and U_{A} is

$$r = \frac{\frac{1}{2(T+1)(T+2)} + \frac{1}{12T}}{\sqrt{\frac{1}{12T}\left(\frac{T}{(T+1)^2(T+2)} + \frac{1}{(T+1)(T+2)} + \frac{1}{12}\right)}}. \tag{5.4}$$

As seen from Eq. (5.4), the correlation between U_{PE} and U_{A} decays as $r \approx T^{-1/2}$ for $T \to \infty$, and practically vanishes in long experiences.

Thus, if Assumptions T1–T2 were correct, rankings of long experiences based on Peak-End total utility would have nothing in common with rankings based on Aver-

[5] It may be of interest to explore how U_{PE} and U_{A} are correlated for stimuli series of a particular pattern.

[6] By doing so, we re-define the stimuli to be the perceived moment utilities themselves. Any other function $g(\cdot)$ would effectively lead to a modification of the distribution of stimuli, which does not qualitatively affect our results.

age total utility. There would be no information redundancy. This follows intuitively from the fact that the information contained in the Peak-End total utility becomes increasingly disconnected from the information contained in the Average total utility as the length of the experience grows. Due to time-separability, no transfer of information is possible across moment utilities.

The vanishingly small correlation coefficient obtained under Assumptions T1–T2 differs dramatically from the correlation observed in the data, which is high and significant even for long experiences (see Table **??**).

5.3.4 Experiences with time non-separable moment utilities and identical individuals

In this section we relax Assumption T1 by allowing moment utilities to depend on the history of experienced stimuli, i.e., generally, $u_t = m(s_1, \ldots, s_{t-1}, s_t)$. To illustrate the implications of this for the correlation between Peak-End and Average total utility, we use the simplest possible form of history dependence: a linear lagged dependent variable model of the form

$$u_t = \eta u_{t-1} + x_t, \quad t = 2, \ldots, T. \tag{5.5}$$

Model (5.5) has been chosen because it nests naturally two prominent processes of subjective judgment. One is the process of anchoring-and-adjustment (Hogarth & Einhorn, 1992), and the other is adaptation (Frederick & Loewenstein, 1999). In Appendix C.2, we describe both models in more detail.

For the anchoring-and-adjustment model, the persistence parameter η in Eq. (5.5) corresponds to parameter $\alpha \in [0, 1]$ in Eq. (C.4), which measures the degree of persistence in moment utilities, while x_t represents weighted transformed stimuli $(1-\alpha)g(s_t)$. For the adaptation model, parameter η represents the speed of adaptation $(1-\beta)$, with $\beta \in [0,1]$, Eq. (C.8), while x_t represents the difference in transformed stimuli, $x_t = g(s_t) - g(s_{t-1})$ (see Appendix C.2).

To examine the implied correlation between Peak-End and Average total utility when moment utilities evolve according to processes described by Eq. (5.5), such as anchoring-and-adjustment or adaptation, we revert to simulations. We find the correlation coefficient r for α and β ranging from 0 to 1 with step size 0.1, and experiences of lengths $T = 2, \ldots, 100$. As before, we assume for simplicity that $g(s) = s$, and the values of s_t are random numbers uniformly distributed between 0 and 1. For every combination (α, T) and (β, T), we generate $N = 10,000$ experiences $(u_{i1}, \ldots u_{iT})$, $i = 1, \ldots, N$, using Eq. (C.4) and Eq. (C.8) with the initial condition $u_0 = 0$. The resulting consistent estimate of r is calculated as the sample correlation coefficient.

Anchoring-and-adjustment is shown to yield correlation r that declines with T, but the value of r becomes systematically larger at higher α's. For $T = 100$ and $\alpha = 0.9$ we obtain $r \approx 0.4$, while without anchoring-and-adjustment [for $\alpha = 0$, Eq. (5.4)] $r \approx 0.12$. Figure 5.1 illustrates this. Figure 5.1 also illustrates the results of simulations for the case of adaptation. Here, correlation between Peak-End and Average total utility tends towards a positive constant and increases with β, the rigidity in adapting. For experiences of length $T = 100$, correlation coefficient $r \approx 0.65$ for $\beta = 0.9$. Weak adaptation leads to an even higher information redundancy than strong anchoring. The reason is that while anchoring-and-adjustment is a pure

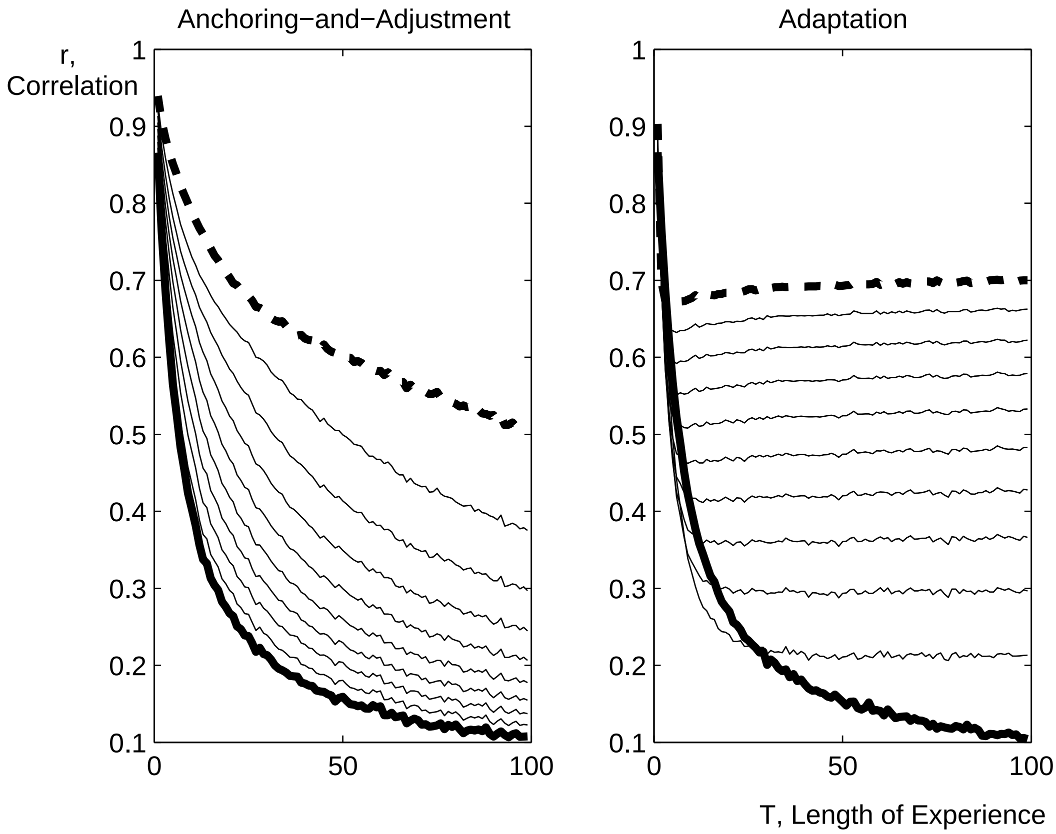


Figure 5.1: Correlation between Peak-End and Average Total Utility in Time Non-Separable Experiences.
Thick dotted lines depict the correlation for experiences characterized by $\alpha = 0.9$ in anchoring-and-adjustment, and $\beta = 1$ in adaptation; thick solid lines depict the correlation for $\alpha = 0$, and $\beta = 0$ (the case of time-separable experiences); lines in-between depict correlations for intermediate values of parameters.

AR(1) process, adaptation, as given by Eq. (C.8), is an ARMA(1,1) process. Serial correlation in terms $g(s_t) - g(s_{t-1})$ of the latter process leads to a more significant information transfer across moments of the experience.

We have shown that even the simplest form of history dependence in perception of stimuli can lead to significant correlation between the Peak-End and Average total utility. However, the values of the correlation coefficient we obtain in simulations are much lower than those observed in the data. To explain the remaining part of the correlation, we have to account for individual heterogeneity, i.e. relax Assumption T2.

5.3.5 Experiences with time non-separable impressions and individual heterogeneity

Individual heterogeneity may enter the dynamics (5.5) of moment utilities in the form of an additive individual-specific term c_i, $i = 1, \ldots, N$, as in panel data models. This term corresponds to the time invariant (average or background) component of the experience that is related to individual characteristics, type of the experience, and initial conditions (the subject's mood prior to the experience). The heterogeneous dynamics then takes the form

$$u_{it} = \eta u_{i,t-1} + x_{it} + c_i. \tag{5.6}$$

Here, unlike in Eq. (5.5), we use subscript i in order to stress the fact that every experience is now individual-specific due to heterogeneity. It is instructive to look at the explicit solution of Eq. (5.6):

$$u_{it} = u_{i1}\eta^{t-1} + \sum_{k=0}^{t-2} \eta^k x_{i,t-k} + c_i \frac{1-\eta^{t-1}}{1-\eta}, \qquad t = 2, \ldots, T. \tag{5.7}$$

The term with the initial condition decays very fast with increasing t, at least for η not too close to 1, and therefore should not contribute much to the total utility, regardless of the aggregation rule. The correlation between Peak-End and Average total utility depends on how much dispersion there is in moment utilities *between* as compared to *within* individual experiences. The dispersion between individuals is determined by the variation of u_{it} across individuals after averaging over time. Such variation comes from the term with c_i in Eq. (5.7), the variance of which is of the order of $\sigma_c^2/(1-\eta)^2$ for large enough T.

The dispersion within a utility profile is determined by the variation of u_{it} over time after averaging across individuals. This variation only comes from the term with transformed stimuli x_{it}. The variance is of the order of $k^2\sigma_s^2/(1-\eta^2)$, where $k = 1-\alpha$ for anchoring-and-adjustment and $k = \sqrt{2}$ for adaptation.[7]

The parameter determining whether the dispersion in individual differences can influence the correlation between Peak-End and Average total utility can be defined as a square root of the ratio of the two variances:

$$v_c = k \left(\frac{1-\eta}{1+\eta}\right)^{1/2} \frac{\sigma_c}{\sigma_s}. \tag{5.8}$$

The correlation between U_{PE} and U_{A} is not affected by individual heterogeneity if $v_c \ll 1$. With $v_c \sim 1$ the role of individual heterogeneity becomes significant.

To explore the role of individual heterogeneity we use simulations with $g(s) = s$, i.i.d uniform stimuli s_{it}, and fixed effects c_i drawn from a distribution with a

[7] In estimating the variances we assumed that $g(s) = s$.

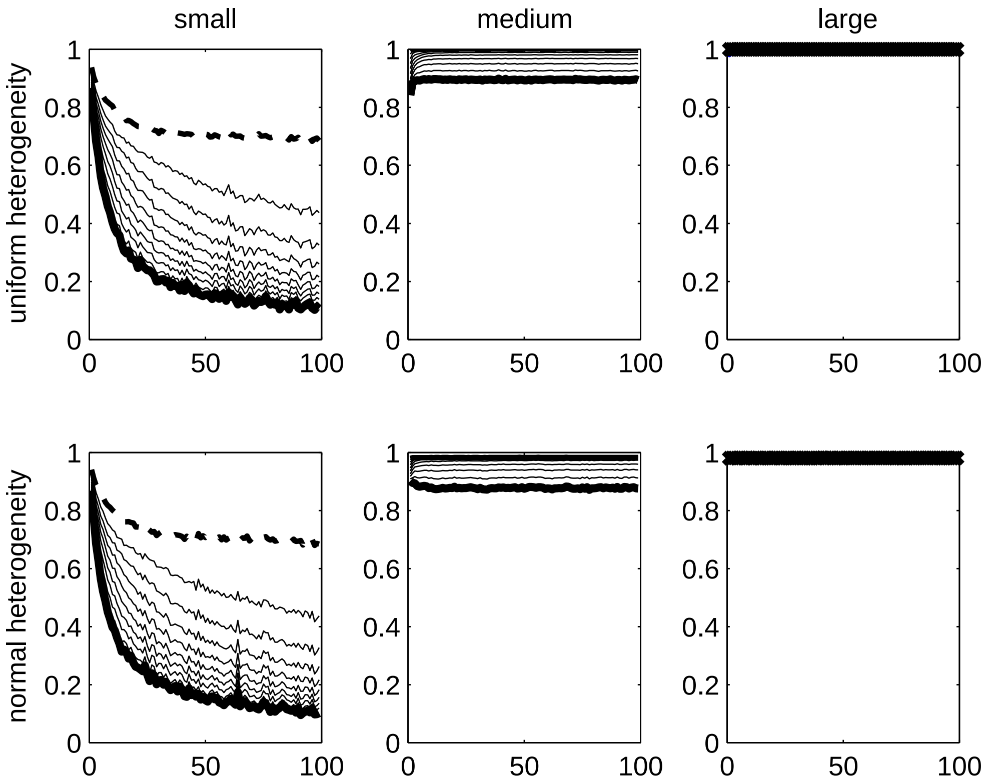


Figure 5.2: Anchoring-and-Adjustment with Individual Heterogeneity.

variance σ_c^2. Figures 5.2–5.3 show the correlation between Peak-End and Average total utility in the presence of *small* ($\sigma_c = 0.2\sigma_s$), *medium* ($\sigma_c = \sigma_s$), and *large* ($\sigma_c = 5\sigma_s$) individual heterogeneity. We used uniform and normal distributions of individual heterogeneity, and both lead to qualitatively similar and expected results: when individual heterogeneity becomes sufficiently dispersed compared to variation of impressions within utility profiles, the correlation is very high.

To summarize, we have shown that individual heterogeneity can further contribute to the redundancy of information about overall evaluations.

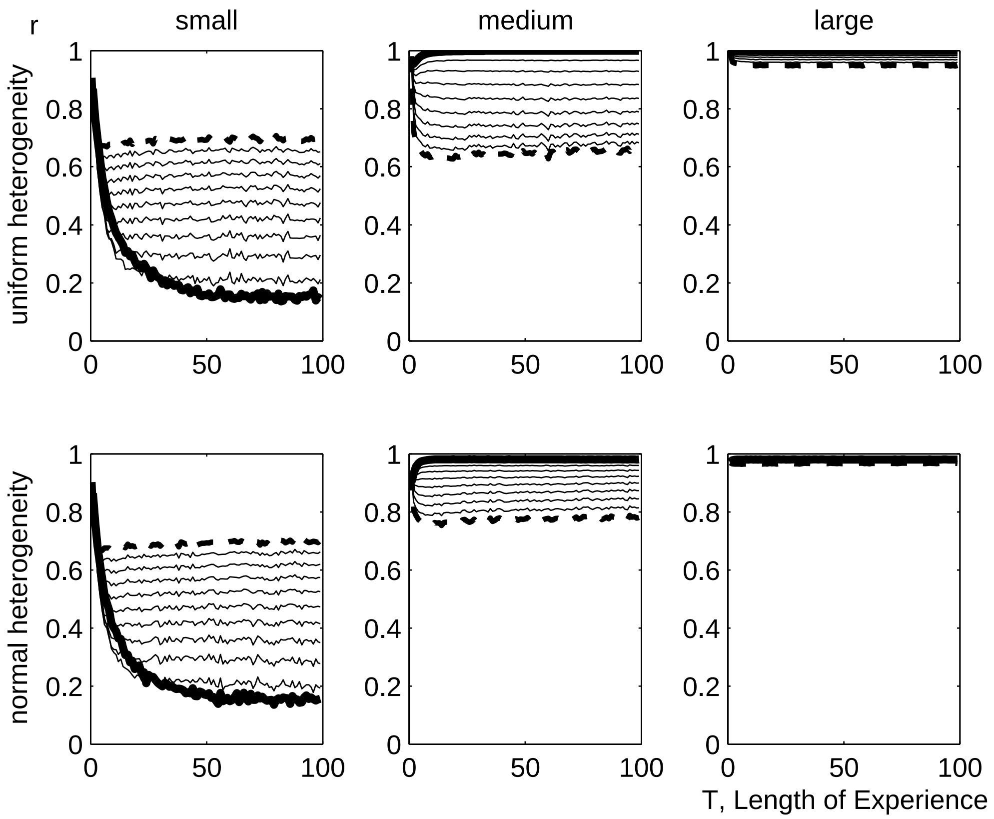


Figure 5.3: Adaptation with Individual Heterogeneity.

5.4 Empirical analysis

5.4.1 Estimation technique

In what follows we estimate the model describing experiences with time non-separable impressions in the data. We control for individual heterogeneity by exploiting the panel structure of available data sets. We consider the following population model:

$$y_{it} = \eta y_{i,t-1} + x_t + c_i + \epsilon_{it}, \quad t = 2, \ldots, T. \tag{5.9}$$

Here, y_{it} is the reported moment utility of subject i after being exposed to an unobserved stimulus s_{it}. We represent the effect of the unobserved stimuli as $g(s_{it}) = x_t + \epsilon_{it}$, where x_t are time-dependent components common to all individuals (to be captured by time dummies), and ϵ_{it} are zero-mean error terms. Note, that, unlike in the previous section, we do not need to assume that x_t are i.i.d. across time. [8] As before, c_i is the unobserved individual-specific effect (to be captured by dummies for individuals). For error terms ϵ_{it}, we make the following standard assumption (Wooldridge, 2002):

Assumption E:

$$\mathrm{E}(\epsilon_{it} | y_{i,t-1}, y_{i,t-2}, \ldots, y_{i0}, x_1, ..., x_T, c_i) = 0. \tag{5.10}$$

According to Assumption E, stimuli are strictly exogenous, while y_{it} is completely determined by its lagged value $y_{i,t-1}$ given the stimuli and the unobserved effect.

[8] In simulations we did not include the error term ϵ_{it} explicitly but used random individual-specific profiles of stimuli s_{it} (by generating a new series of stimuli for every individual). By doing so we effectively introduced zero-mean error terms $\epsilon_{it} = \alpha(s_{it} - s_t)$ for anchoring-and-adjustment and $\epsilon_{it} = s_{it} - s_t - (s_{i,t-1} - s_{t-1})$ for adaptation. The values s_t then can be interpreted as s_{it} averaged across individuals.

We estimate model (5.9) by implementing the Arellano-Bond panel data estimator (Arellano, 2003). It involves first-differencing Eq. (5.9) to remove the unobserved effect, and using the GMM estimation. We use $y_{i,t-2}$ and $y_{i,t-3}$ as instruments. Given the parameter estimate $\hat{\eta}$, we then recover the underlying time-specific stimuli, $\hat{x}_t$, and the individual fixed effects, $\hat{c}_i$, in a regression of a new variable $z_{it} = (y_{it} - \hat{\eta} y_{i,t-1})$ on time dummies and individual dummies [see Eq. (5.9)].

We use estimation results to predict the correlation between Peak-End and Average total utility. In order to do that, we kernel-estimate the distribution of fitted transformed stimuli $\hat{x}_t$ and fitted fixed effects $\hat{c}_i$, and recover their means and variances, $\hat{\sigma_s^2}$ and $\hat{\sigma_c^2}$, in the data. We also recover the variance of residuals $\hat{\epsilon}_{it}$. Then for each data set we use its length of the experience T, the estimated persistence coefficient $\hat{\eta}$, and means and variances of $\hat{x}_t$, $\hat{c}_i$, and $\hat{\epsilon}_{it}$ to simulate experiences. In these simulations we assume normality (which cannot be rejected in most cases, see below). The experiences simulated are summarized by Peak-End rule and the Average rule. Finally, we examine the correlation between Peak-End and Average total utility for each simulated data set, and compare it to the empirical correlation in the original data set. This shows the extent to which accounting for the nature of moment utilities helps predict the redundancy of information about overall evaluations of experiences.

5.4.2 Estimation results

Estimation results are presented in Table 5.2. These refer to 51 data sets for which the number of observations was sufficient to serve for estimation. [9]

[9] In three data sets, 37, 38 and 42, the number of observations was not sufficient to serve for estimation.

Table 5.2: Estimation Results for the Persistence in Moment Utilities ($\hat{\eta}_{AB}$).

Data Set	T, Experience Length	N, Participants	$\hat{\eta}_{AB}$,	95% Conf. Interval
1	30	26	0.56	0.46 – 0.66
2	30	27	0.81	0.72 – 0.89
3	30	27	0.87	0.78 – 0.95
4	30	27	0.33	-0.05 – 0.71
5	30	27	0.61	0.25 – 0.97
6	30	27	0.39	0.17 – 0.62
7	30	27	0.88	0.80 – 0.96
8	30	27	0.62	0.54 – 0.71
9	30	27	0.76	0.64 – 0.87
10	30	27	0.58	0.20 – 0.96
11	42	23	0.81	0.62 – 0.99
12	45	27	0.72	0.52 – 0.92
13	45	27	0.72	0.66 – 0.78
14	45	26	0.81	0.67 – 0.96
15	50	27	0.72	0.55 – 0.89
16	50	27	0.92	0.89 – 0.95
17[a]	52	24	0.90	0.79 – 1.01
18[a]	57	22	0.76	0.67 – 0.86
19	60	26	0.51	0.31 – 0.71
20	60	27	0.44	0.19 – 0.69
21	60	26	0.71	0.59 – 0.83
22	60	27	0.71	0.50 – 0.92
23	60	27	0.89	0.84 – 0.94
24	60	27	0.87	0.78 – 0.95
25	60	27	0.72	0.51 – 0.94

[a]Average autocovariance in residuals of order 2 is not 0.

Table 2 (continued).
Estimation Results for the Persistence in Moment Utilities ($\hat{\eta}_{AB}$).

Data Set	T, Experience Length	N, Participants	$\hat{\eta}_{AB}$,	95% Conf. Interval
26	60	27	0.82	0.77 – 0.87
27	60	27	0.76	0.66 – 0.86
28	60	27	0.74	0.59 – 0.89
29	73	25	0.62	0.54 – 0.70
30	75	27	0.65	0.52 – 0.79
31	90	26	0.76	0.62 – 0.90
32	90	27	0.75	0.63 – 0.87
33[a]	90	27	0.91	0.84 – 0.97
34	90	26	0.77	0.69 – 0.85
35	30	23	0.21	0.14 – 0.29
36	30	23	0.23	0.12 – 0.34
39[a]	18	42	0.33	0.15 – 0.52
40	7	36	0.03	-0.18 – 0.24
41	7	46	0.20	-0.01 – 0.42
43	6	28	-0.11	-0.31 – 0.09
44	11	37	-0.05	-0.23 – 0.14
45	10	34	0.06	-0.17 – 0.30
46	10	35	0.17	-0.02 – 0.36
47	10	35	0.07	-0.16 – 0.31
48	10	35	-0.19	-0.44 – 0.05
49	10	35	0.08	-0.14 – 0.30
50	10	35	0.00	-0.18 – 0.18
51	10	35	0.12	-0.15 – 0.39
52	10	35	0.15	0.01 – 0.29
53	10	35	0.16	-0.04 – 0.37
54[a]	10	35	0.25	0.07 – 0.44

[a]Average autocovariance in residuals of order 2 is not 0.

In 38 data sets, the estimate $\hat{\eta}$ is statistically significant (for those data sets the 95% confidence intervals do not contain zero), its value ranges from 0.15 to 0.92. Recall that the value of $\hat{\eta}$ corresponds to the value of α if moment utilities evolve according to anchoring-and-adjustment, and it corresponds to $(1-\beta)$ if moment utilities follow the process of adaptation.

Estimated individual-specific effects $\hat{c}_i$ provide information about the distribution of individual heterogeneity in the data. Normality tests based on skewness and kurtosis, as well as overall tests of normality, suggest that in 34 of 51 data sets normality cannot be rejected at 5% significance level. We are, therefore, able to characterize individual heterogeneity for each data set in terms of a normal distribution with the mean and standard deviation equal to the sample mean and standard deviation of $\hat{c}_i$ in the corresponding data set.

Estimated stimuli, $\hat{x}_t$ are also assumed to follow a normal distribution. Normality cannot be rejected for 17 out of 36 data sets with 30 and more moments of experience. Recall that $\hat{x}_t$ proxy the scaled stimuli $(1-\alpha)g(s_t)$ if moment utilities evolve under anchoring-and-adjustment and the difference in stimuli $g(s_t) - g(s_{t-1})$ if there is adaptation.[10]

5.4.3 Explaining information redundancy

Table 5.1, column 6, reports the predicted correlation between Peak-End and Average total utility, $r_{\text{predicted}}$, for each data set. Our prediction is based on the sim-

[10]In our illustrative simulations in Section *The Model* we have assumed a uniform distribution for s_t, a non-crucial assumption since function $g(\cdot)$ is unknown, and the effect of the stimuli has to be captured by time dummies.

ulations of experiences given estimation results for each data set. We assumed that moment utilities in the data set evolved with the persistence parameter $\hat{\eta}$, and took into account the mean and the standard deviation of estimated individual heterogeneity, stimuli, and the error term.

The prediction of the correlation between Peak-End and Average total utility is rather accurate. Its mean absolute deviation from the correlation observed in the data is 0.09. Moreover, it captures well the variation in the correlation between Peak-End and Average total utility across data sets. The correlation between $r_{\text{predicted}}$ and the correlation observed in the data is high ($r(51) = 0.77$, $p < 0.001$, and $r_{Spearman}(51) = 0.72$, $p < 0.001$). Importantly, this prediction is rather general. We used distributional properties, and not the actual estimates of individual heterogeneity, stimuli, and the error. [11]

5.5 Discussion

We have analyzed the redundancy of information about overall evaluations of experiences as the reason for not learning the true composition of overall evaluations. Previous research has examined whether Peak-End or average impressions predicted better overall evaluations of experiences. Some findings did not point to the exclusivity of a single rule (Kahneman, 2000), but other results suggested that the Peak-End rule was a better predictor of overall evaluations and no control for average impressions was necessary (Fredrickson & Kahneman, 1993; Redelmeier & Kahneman, 1996;

[11] Prediction based on the estimates of individual-specific effects and time-specific stimuli is more accurate (mean absolute deviation is 0.04 and correlation between r_{data} and $r_{predicted}$ equal to 0.87.)

Ariely & Carmon, 2000). To our knowledge, nobody has focused on the correlation between the two alternatives, and studied it systematically. We have done so in this work as a measure of information redundancy in the Peak-End rule learning environment.

Intuitively, the two rules, U_{PE} and U_{A}, are dissimilar, especially if applied to long experiences. And yet, in 54 data sets on experiences lived across time, we have found a high correlation between Peak-End and Average impressions. Our contribution has been to identify conditions, conducive to high information redundancy. These include the following:

(a) Stimuli are not perceived at their face value; impression at time t contains information about impressions at previous moments, $t-1$, $t-2$, etc., due to psychological processes, such as the anchoring-and-adjustment or adaptation (other psychological processes that can make impressions time non-separable may have similar effects).

(b) Impressions contain individual-specific effects c_i. These can raise the dispersion of impressions *between* individuals above the dispersion in impressions *within* individuals and thereby explain the observed high correlation between U_{A} and U_{PE}.

In validating these ideas empirically, we have estimated the model of time non-separable impressions with individual heterogeneity using panel data on experiences lived across time. The model nests two prominent psychological processes, anchoring-and-adjustment and adaptation, which are responsible for the transfer of information across the moments of the experience. It turned out, persistence in most data sets exceeded 0.50. Estimation results were shown to help explain the magnitude of correlation between Peak-End and Average impressions and its variation across data sets.

The factors contributing to the redundancy of information about overall evaluations can explain why the notions of "overall" and "average" are treated as equivalent. To develop a belief in the Peak-End rule, people have to be sensitive to an often small difference in the ability of Peak-End and average impressions to predict overall evaluations. Moreover, this has to happen given a commonly observed high correlation between the two alternatives, inherent in the psychological process of experiencing. Laboratories and a sophisticated analytical toolbox can be helpful. Otherwise, it is, certainly, not easy to learn the differential weights that impressions have for the overall evaluation. Thus, the belief that overall evaluations are due to average impressions, or, lack of differentiation, may prevail.

5.5.1 Implications for theory

We have presently shown how the specific psychological processes by which moment utilities evolve in a given experience can make Peak-End utility contain sufficient information about the average of all moment utilities. We have, thus, explored one implication of the specific form of utility *inclusiveness* (Kahneman et al., 1997).

Our work is related to the explorations of unit weighting schemes for decision making. Einhorn and Hogarth (1975) examine the general problem of forming composite variables from components. Single components are compared to the best linear combination of all components in terms of their ability to predict the composite variable. The success of the prediction based on a single component is found to increase with intercorrelation of the components. Our work has been informative about simple one-parameter dynamic processes inducing a particular structure of such intercorrelation (moment utilities in our analysis were the "components"). Structures of intercorrelation induced have been shown to relate single components to the average of all components. We have explored how the process parameter affects the strength of this relationship.

5.5.2 Applications

In economics, there has been an increasing use of data on subjective well-being (Kahneman & Krueger, 2006). The decision of how to summarize a history of subjective well-being or concerns about what well-being judgments reflect are important to economists. Concepts of "impressions", "customer satisfaction", "job satisfaction" and "performance appraisals" in psychology, marketing and organizational behavior, raise similar issues. In these fields, there is a tradition of studying subjective judgments as either dependent or independent variables, and relating them to a future choice of experiences, product purchase, managerial decisions of reward or punishment, and work-related behaviors (Wirtz et al., 2003; Oishi & Sullivan, 2006; DeNisi, 1996; Lam, Dalal, Weiss, & Welch, 2006). Finally, practitioners often take decisions based on the measurement of subjective judgments. The measurements are frequently documented, and the need to aggregate multiple measurements across time is natural. Suppose, you wanted to know who of your customers is more happy with a new service, or, who of your employees is less satisfied with the job. Although driven naturally to average across all measurements, you may follow the literature on overall evaluations, and choose to use the Peak-End rule instead. Our work is informative about the conditions under which the choice of the aggregation rule is or is not likely to be crucial for your judgment.

References

Alexandrova, A. (2005). Subjective well-being and Kahneman's "objective happiness". *Journal of Happiness Studies*, *6*, 301-324.

Arellano, M. (2003). *Panel data econometrics.* Oxford University Press.

Ariely, D. (1998). Combining experiences over time: The effects of duration, intensity changes and on-papers line measurements on retrospective pain evaluations. *Journal of Behavioral Decision Making*, *11*, 19-45.

Ariely, D., & Carmon, Z. (2000). Gestalt characteristics of experiences: The defining features of summarized events. *Journal of Behavioral Decision Making*, *13*, 191-201.

Ariely, D., & Carmon, Z. (2003). The sum reflects only some of its parts: A critical overview of research about summary assessment of experiences. In D. R. G. Loewenstein & R. Baumeister (Eds.), *Time and decision: Economic and psychological perspectives on inter-temporal choice* (p. 323-350). New York: Russel Sage Foundation.

Ariely, D., & Loewenstein, G. (2000). The importance of duration in ratings of, and choices between, sequences of outcomes. *Journal of Experimental Psychology: General*, *129*, 508-523.

Ariely, D., Loewenstein, G., & Prelec, D. (2003). Coherent arbitrariness: Stable demand curves without stable preferences. *Quarterly Journal of Economics*, *118*, 73-105.

Ariely, D., Loewenstein, G., & Prelec, D. (in press). Tom Sawyer and the construction of value. *Journal of Economic Behavior and Organization*, *XX*, XXXX.

Ariely, D., & Zauberman, G. (2000). On the making of an experience: The effects of breaking and combining experiences on their overall evaluation. *Journal of Behavioral Decision Making*, *13*, 219-232.

Ariely, D., & Zauberman, G. (2003). Differential partitioning of extended experiences. *Organizational Behavior and Human Decision Processes*, *91*, 128-139.

Bagozzi, R., Gopinath, M., & Nyer, P. (1999). The role of emotions in marketing. *Journal of the Academy of Marketing Science*, *27*, 184-206.

Baumgartner, H., Sujan, M., & Padgett, D. (1997). Patterns of affective reactions to advertisements: The integration of moment-to-moment responses into overall judgments. *Journal of Marketing Research*, *XXXIV*, 219-232.

Bradley, M., & Lang, P. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25(1)*, 49-59.

Brendl, M., & Higgins, T. (1995). Principles of judging valence: What makes events positive or negative? In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. vol.28, 95-160). New York: Academic Press.

Carmon, Z., & Kahneman, D. (1996). The experienced utility of queuing: Experience profiles and retrospective evaluations of simulated queues. *working paper*, *Fuqua School of Business*, Duke University.

Cojuharenco, I. (2007). Lay intuitions about overall evaluations of experiences. *Judgment and Decision Making*, *2 (1)*, 40-47.

DeNisi, A. (1996). *A cognitive approach to performance appraisal: A program of research.* London: Routledge.

Diener, E., Wirtz, D., & Oishi, S. (2001). End effects of rated life quality: The James Dean effect. *Psychological Science*, *12*, 124-128.

Di Tella, R., & McCulloch, R. (2006). Some uses of happiness data in economics. *Journal of Economic Perspectives*, *20*, 25-46.

Einhorn, H., & Hogarth, R. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance*, *13*, 171-192.

Forgas, J. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin*, *117(1)*, 39-66.

Forgas, J., & George, J. (2001). Affective influences on judgments and behavior in organizations: An information processing perspective. *Organizational Behavior and Human Decision Processes*, *86(1)*, 3-34.

Frederick, S., & Loewenstein, G. (1999). Hedonic adaptation. In D. Kahneman, E. Diener, & N. Schwartz (Eds.), *Well-being: The foundations of hedonic psychology* (p. 302-330). New York: Russel Sage Foundation.

Fredrickson, B. (2000). Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition and Emotion*, *14*, 577-606.

Fredrickson, B., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology*, *65*, 45-55.

Gardner, M. (1985). Mood states and consumer behavior: A critical review. *Journal of Consumer Research*, *12(3)*, 281-300.

Gnoth, J., Zins, A., Lengmueller, R., & Boshoff, C. (1999). Dimensions of emotions, mood and motivational orientation with regard to experiential consumption. In L. Hindebrandt, D. Annacker, & D. Klapper (Eds.), *Proceedings of the 27th emac conference.* Berlin: Humboldt University.

Hands, D., & Avons, S. (2001). Recency and duration neglect in subjective assessment of television picture quality. *Applied Cognitive Psychology*, *15*, 639-657.

Hogarth, R., & Einhorn, H. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, *24*, 1-55.

Hsee, C., & Abelson, R. (1991). The velocity relation: Satisfaction as a function of the first derivative of outcome over time. *Journal of Personality and Social Psychology*, *60*, 341-347.

Huber, O., Wider, R., & Huber, O. (1997). Active information search and complete information presentation in naturalistic risky decision tasks. *Acta Psychologica*, *95*, 15-29.

Hunter, J., Schmidt, F., & Jackson, G. (1982). *Meta-analysis: Cumulating research findings across studies.* Sage publications.

Kahneman, D. (2000). Evaluation by moments: Past and future. In D. Kahneman & A. Tversky (Eds.), *Choices, values and frames* (p. 693-708). Cambridge University Press.

Kahneman, D., Fredrickson, D., Schreiber, C., & Redelmeier, D. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science*, *4*, 401-405.

Kahneman, D., & Krueger, A. (2006). Developments in the measurement of subjective well-being. *Journal of Economic Perspectives*, *20*, 3-24.

Kahneman, D., Wakker, P., & Sarin, R. (1997). Back to bentham? explorations of experienced utility. *The Quarterly Journal of Economics*, 375-405.

Karelaia, N., & Hogarth, R. (2007). Determinants of linear judgment: A meta-analysis of lens model studies. *Technical Report. Universitat Pompeu Fabra.*

Lam, H., Dalal, R., Weiss, H., & Welch, E. (2006). Momentary versus retrospective reports of mood, behavior, and performance. *Presented at the annual meeting of the Society for Industrial and Organizational Psychology (Dallas, TX).*

Lang, P., Bradley, M., & Cuthbert, B. (2005). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. *Technical Report A-6. University of Florida*, *25(1)*, Gainesville, FL.

Langer, T., Sarin, R., & Weber, M. (2005). The retrospective evaluation of payment sequences: Duration neglect and peak-and-end effects. *Journal of Economic Behavior and Organization*, *58*, 157-175.

Loewenstein, G. (1987). Anticipation and the valuation of delayed consumption. *The Economic Journal*, *97*, 666-684.

Loewenstein, G., & Prelec, D. (1993). Preferences for sequences of outcomes. *Psychological Review*, *100*, 91-108.

Luomala, H., & Laaksonen, M. (2000). Contributions from mood research. *Psychology and Marketing*, *17(3)*, 195-232.

Manis, M. (1971). Context effects in communication: Determinants of verbal output and referential decoding. In M. H. Appley (Ed.), *Adaptation-level theory* (p. 237-255). New York: Academic Press, Inc.

Oishi, S., & Sullivan, H. (2006). The predictive value of daily vs. retrospective well-being judgments in relationship stability. *Journal of Experimental Social Psychology*, *42*, 460-470.

Pieters, R., & Van Raaij, F. (1988). Functions and management of affect: Applications to economic behavior. *Journal of Economic Psychology*, *9*, 251-282.

Read, D., Antonides, G., Ouden, L. van den, & Trienekens, H. (2001). Which is better: Simultaneous or sequential choice? *Organizational Behavior and Human Decision Processes*, *84(1)*, 54-70.

Read, D., Loewenstein, G., & Rabin, M. (1999). Choice bracketing. *Journal of Risk and Uncertainty*, *19*, 171-197.

Redelmeier, D., & Kahneman, D. (1996). Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain*, *66*, 3-8.

Robbins, T., & DeNisi, A. (1998). Mood versus interpersonal affect: Identifying process and rating distortions in performance appraisal. *Journal of Business and Psychology*, *12(3)*, 313-325.

Robinson, M., & Clore, G. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, *128*, 934-960.

Rozin, A., Rozin, P., & Goldberg, E. (2004). The feeling of music past: How listeners remember musical affect. *Music Perception*, *22*, 15-39.

Schreiber, C., & Kahneman, D. (2000). Determinants of the remembered utility of aversive sounds. *Journal of Experimental Psychology: General*, *129*, 27-42.

Schwarz, N., & Clore, G. (2003). Mood as information: 20 years later. *Psychological Inquiry*, *14*, 296-303.

Schwarz, N., & Clore, G. (2006). Feelings and phenomenal experiences. In E. Higgins & A. Kruglanski (Eds.), *Social psychology: A handbook of basic principles* (p. xxx). New York: Guilford Press.

Stone, A., Shiffman, S., & DeVries, M. (1999). Ecological momentary assessment. In D. Kahneman, E. Diener, & N. Schwartz (Eds.), *Well-being: The foundations of hedonic psychology* (p. 26-39). New York: Russel Sage Foundation.

Treisman, A., & Chong, S. (2003). Representation of statistical properties. *Vision Research*, *43*, 393-404.

Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review*, *110*, 403-421.

Updegraff, J., Gable, S., & Taylor, S. (2004). What makes experiences satisfying? The interaction of approach-avoidance motivations and emotions in well-being. *Journal of Personality and Social Psychology*, *86*, 496-504.

Varey, C., & Kahneman, D. (1992). Experiences extended across time: Evaluation of moments and episodes. *Journal of Behavioral Decision Making*, *5*, 169-186.

Wegener, D., Petty, R., & Smith, S. (1995). Positive mood can increase or decrease message scrutiny: The hedonic contingency view of mood and message processing. *Journal of Personality and Social Psychology*, *69(1)*, 5-15.

Wirtz, D., Kruger, J., Napa Scollon, C., & Diener, E. (2003). What to do on spring break? The role of predicted, on-line and remembered experience in future choice. *Psychological Science*, *14*, 520-524.

Wooldridge, J. (2002). *Econometric analysis of cross section and panel data.* MIT press.

Appendix A

Appendix to Chapter 2

A.1 Study 1 sample response sheets (explanations and overall)

Please, indicate your identification letter here : H

Problem Set 6.

Problem 1. Ronald McDonald.

a. 0 1 2 3 4 5 6 7 8 9 10

b. 0 1 2 3 4 5 6 7 8 9 10

c. 0 1 2 3 4 5 6 7 8 9 10

d. 0 1 2 3 4 5 6 7 8 9 10

e. 0 1 2 3 4 5 6 7 8 9 10

f. 0 1 2 3 4 5 6 7 8 9 10

g. 0 1 2 3 4 5 6 7 8 9 10

h. 0 1 2 3 4 5 6 7 8 9 10

i. 0 1 2 3 4 5 6 7 8 9 10

j. 0 1 2 3 4 5 6 7 8 9 10

Remember, you are evaluating how ***engaging*** the solution to the particular question is, with "0" meaning "not engaging at all" and "10" meaning "extremely engaging".

Problem 2. Tricky Dick.

a. 0 1 2 3 4 5 6 7 8 9 10

b. 0 1 2 3 4 5 6 7 8 9 10

c. 0 1 2 3 4 5 6 7 8 9 10

c. 0 1 2 3 4 5 6 7 8 9 10

d. 0 1 2 3 4 5 6 7 8 9 10

e. 0 1 2 3 4 5 6 7 8 9 10

f. 0 1 2 3 4 5 6 7 8 9 10

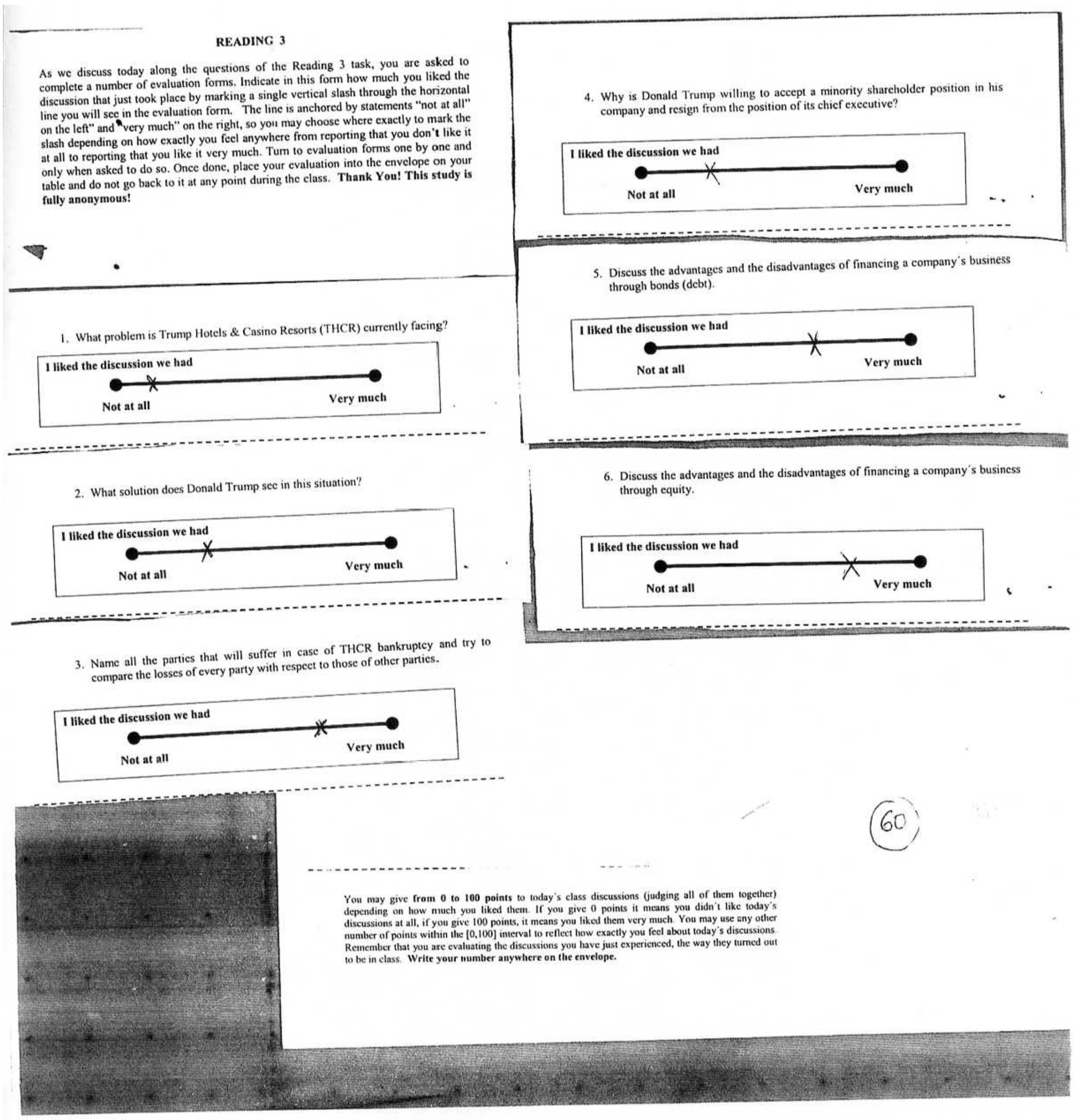
g. 0 1 2 3 4 5 6 7 8 9 10

Please, indicate your identification letter here: H

Overall, the totality of solutions presented in the practice session was as engaging as:

0 1 2 3 4 5 6 7 8 9 10

A.2 Study 2 sample response forms (discussions and overall)



READING 3

As we discuss today along the questions of the Reading 3 task, you are asked to complete a number of evaluation forms. Indicate in this form how much you liked the discussion that just took place by marking a single vertical slash through the horizontal line you will see in the evaluation form. The line is anchored by statements "not at all" on the left" and "very much" on the right, so you may choose where exactly to mark the slash depending on how exactly you feel anywhere from reporting that you don't like it at all to reporting that you like it very much. Turn to evaluation forms one by one and only when asked to do so. Once done, place your evaluation into the envelope on your table and do not go back to it at any point during the class. **Thank You! This study is fully anonymous!**

1. What problem is Trump Hotels & Casino Resorts (THCR) currently facing?

I liked the discussion we had

Not at all **Very much**

2. What solution does Donald Trump see in this situation?

I liked the discussion we had

Not at all **Very much**

3. Name all the parties that will suffer in case of THCR bankruptcy and try to compare the losses of every party with respect to those of other parties.

I liked the discussion we had

Not at all **Very much**

4. Why is Donald Trump willing to accept a minority shareholder position in his company and resign from the position of its chief executive?

I liked the discussion we had

Not at all **Very much**

5. Discuss the advantages and the disadvantages of financing a company's business through bonds (debt).

I liked the discussion we had

Not at all **Very much**

6. Discuss the advantages and the disadvantages of financing a company's business through equity.

I liked the discussion we had

Not at all **Very much**

60

You may give **from 0 to 100 points** to today's class discussions (judging all of them together) depending on how much you liked them. If you give 0 points it means you didn't like today's discussions at all, if you give 100 points, it means you liked them very much. You may use any other number of points within the [0,100] interval to reflect how exactly you feel about today's discussions. Remember that you are evaluating the discussions you have just experienced, the way they turned out to be in class. **Write your number anywhere on the envelope.**

A.3 Study 3 sample response forms

A.3.1 Weekly

A	N	I	V

*

1. The list below contains suggested readings for this week (February 14-15, 2005). Please, check the box(es) of the reading(s) you did and indicate how much you enjoyed each paper read (0-100 points). Only evaluate readings you did <u>in addition to</u> previously read material (you evaluated other readings already in the questionnaire last week).

☒	Reading	Points
	18. Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin*, 107, 296-310.	70
	19. Slovic, P. (1991). The construction of preferences. *American Psychologist*, 50(5), 364- 71.	70
	20. Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology*, 60 (2), 181-192. ***	90
	21. Yaniv, I., & Hogarth, R. M. (1993). Judgmental versus statistical prediction: Information asymmetry and combination rules. *Psychological Science*, 4 (1), 58-62.***	85

2. Did you do any other readings for <u>this</u> week which do not appear in the table above? Please, evaluate them here. → [ ]

3. Express your general impression about <u>all the readings</u> done for this week (judging all of them together). → 85

* Please, write down a combination of 4 letters, where the first 2 letters would be the first 2 letters of your first name and the second two letters – the first 2 letters of your last name.

A.3.2 Overall

A	N	I	V

1. The list below contains suggested readings for this week (March 14-15, 2005). Please, check the box(es) of the reading(s) you did and indicate how much you enjoyed each paper read (0-100 points). Only evaluate readings you did <u>in addition to</u> previously read material (you evaluated other readings already in the questionnaire last week).

☒	Reading	Points
	34. Camerer, C. F., Babcock, L., Loewenstein, G., & Thaler, R. H. (1997). Labor supply of New York City cab drivers: One day at a time. *The Quarterly Journal of Economics*, 112 (2), 407-441.	100
	35. Odean, T. (1998). Are investors reluctant to realize their losses? *Journal of Finance*, 53(5), 1775-1798.	70
	36. Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive consumer choice processes. *Journal of Consumer Research*, 25, 187-217.	90
	37. Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral Decision Making*, 12, 183-206.	89
	38. Arkes, H., & Ayton, P. (1999). The sunk cost and Concorde effects: Are humans less rational than lower animals. *Psychological Bulletin*, 125 (5), 591-600.	95

2. Did you do any other readings for <u>this</u> week which do not appear in the table above? Please, evaluate them here.

→ 50

3. Express your general impression about all the readings done for <u>this week</u> (judging all of them together).

→ 95

4. How much did you enjoy <u>all the readings</u> you have done for this course (judging all the readings you have done for this class together)? Mark a single vertical slash through the horizontal line below:

Not at all — Very much

5. How would you express your overall evaluation of the readings for this course in points, from 0 to 100?

→ 95

6. Please, explain briefly the reasons for your overall evaluation of the readings for this course?

I have enjoyed the readings for the course, they gave me an opportunity to rethink many issues related to human's behavior and reasons which drive that ~~behavior~~ behavior.
- Very useful for scientists as well as for managers.

7. Can you provide an estimate of average weekly rating you have given to course readings in previous questionnaires?

→ 85

THANK YOU FOR PARTICIPATING IN THIS FIELD STUDY!

Appendix B

Appendix to Chapter 3

B.1 Post-experimental questionnaire

CUESTIONARIO

POR FAVOR, RELLENA EL CUESTIONARIO PÁGINA POR PÁGINA EN EL ORDEN ESTABLECIDO Y NO CAMBIES TUS RESPUESTAS.

1. En la barra de abajo indica como ha sido tu experiencia con las imágenes que acabas de ver, marcando una línea vertical sobre la barra.

He vivido una experiencia absolutamente horrible | He vivido una experiencia neutral | He vivido una experiencia totalmente agradable

2. Evaluando en puntos, de –100 a 0 y a 100, ¿cuánto te ha agradado o desagradado tu experiencia? Indica tu grado de desagrado utilizando cualquier número de puntos de 0 a –100, y tu agrado de 0 a 100.

3. ¿Cuantas ganas tienes de repetir la misma experiencia? Marca la casilla que se corresponde con tu opinión.

1	2	3	4	5	6	7
No la repetiría ni aunque me lo pagaran	La repetiría si me lo pagaran bien	La repetiría si me pagaran algo	No tengo ganas de repetirla	Tengo ganas a repetirla pero no pagaría nada para hacerlo	Tengo ganas y pagaría algo para hacerlo	Tengo muchísimas ganas y pagaría mucho para hacerlo

4. Ahora, por favor evalúa la serie de imágenes que acabas de ver, utilizando puntos de –100 (la secuencia ha sido totalmente desagradable) a 100 (la secuencia ha sido totalmente agradable), donde 0 puntos significa que no ha sido ni agradable, ni desagradable. Explica, por favor, ¿qué ha determinado tu opinión? ¿En qué criterios te has basado para evaluar la serie?

5. Para recordar tu impresión general de la serie, ¿qué imagen consultarías? Descríbela.

<u>Verifica que hayas contestado a todas las preguntas y pasa a la página siguiente. ¡Gracias!</u>

¿Cómo te sientes ahora? Marca una figura en cada fila abajo que mejor represente lo que sientas (la primera fila de arriba describe la dirección de tus sentimientos: de positivos a negativos, la segunda – la intensidad: de fuerte a débil, y la tercera – tu capacidad de controlar la situación: de sentirte sin control a sentirte importante, completamente en control). También puedes marcar los espacios entre las figuras para mayor precisión.

[SAM INSTRUMENT HERE]

6. Volvamos a la serie de imágenes que acabas de ver. ¿Podrías estimar la media de las evaluaciones dadas a cada imagen? Indicala, por favor.

7. ¿Cómo era la primera imagen de la secuencia? ¿Que contenía?

8. ¿Cómo era la última imagen de la secuencia? ¿Que contenía?

9. ¿Cómo era la imagen que más te ha agradado? ¿Que contenía?

10. Qué imagen elegirías como la más representativa de la serie? Descríbela.

Verifica que hayas contestado a todas las preguntas y pasa a la página siguiente. ¡Gracias!

11. ¿Cómo era la imagen que menos te ha agradado? ¿Que contenía?

12. ¿Que imagen ha sido la más relevante para ti? ¿Cuál era el contenido?

13. Recuerda, por favor, los diferentes contenidos de las imágenes y haz una lista exhaustiva de ellos.

14. Imagina que tiene que estimar la evaluación de la experiencia de otra persona con una serie de imágenes y puedes hacerle 2 preguntas para hacer tu estimación. Solo sabrás que la evaluación ha sido un número de puntos entre 0 y 100, que la serie fue de diferente cantidad y contenido de imágenes que la tuya, pero que no había ninguna imagen desagradable**,** y que la persona ha evaluado cada imagen de la serie utilizando la misma escala de 0 a 100 con las imágenes más agradables recibiendo más puntos. ¿Que le preguntarías?

1.

2.

15. Imagina las respuestas a tus preguntas y explica como estimarías la evaluación de la serie basándote en ellas.

Verifica que hayas contestado a todas las preguntas y pasa a la página siguiente. ¡Gracias!

16. Imagina que para estimar la evaluación de la serie de imágenes vista por la otra persona, solo tendrías el derecho de hacer 2 preguntas de la lista de abajo. Tu estimación de la evaluación de la serie se basará sobre las respuestas a las dos preguntas elegidas. ¿Que preguntas elegirías?

- ¿Cuál fue la evaluación de la imagen más desagradable?
- ¿Dónde en la serie se encontraba tu imagen favorita?
- ¿Cuál fue tu evaluación de la imagen que estaba en el medio de la serie?
- ¿Cuál fue tu evaluación de la primera imagen de la serie?
- ¿Cuál fue la media de tus evaluaciones de imágenes en esta serie?
- ¿Cuántas imágenes componían la serie?
- ¿Cuál fue la evaluación de la última imagen?
- ¿Cuál fue la evaluación de tu imagen favorita?
- ¿Cuál fue la suma de las evaluaciones de las imágenes en esta serie?
- ¿Cómo fue la progresión de las imagines en esta serie: de desagradable a agradable o al revés?
- ¿Cuál fue la evaluación mas frecuente que has utilizado en evaluar esta serie de imágenes?
- ¿Fueron las evaluaciones de las imágenes muy diferentes o muy similares en esta serie?

17. ¿Que pregunta elegirías si solo podrías elegir **una**? Subráyala.

18. ¿Y que dos preguntas elegirías si la serie consistiera de imágenes desagradables y ninguna imagen agradable? Marca estas dos preguntas con el símbolo -- y cruza con una sola línea la pregunta que elegirías si solo pudieras elegir una ~~(abcdef..).~~

19. ¿Te sentías estresado por la restricción del tiempo en evaluar las imágenes? Si No Un poco

Indica, por favor, tus datos:

NOMBRE: E-MAIL:

SEXO: F/M EDAD: ESTUDIANTE DE _______________________

Levanta tu mano para entregar este cuestionario. ¡Gracias!

Appendix C

Appendix to Chapter 5

C.1 Deriving correlation between Peak-End and Average total utility when $u_t = s_t$, and s_t are i.i.d. uniform

Without loss of generality, the stimuli can be enumerated so that $s_1 \geq s_2 \geq \ldots \geq s_T$. The joint probability density of the stimuli then can be written as

$$f(s_1, \ldots, s_T) = T! I_{[0,1]}(s_1) \ldots I_{[0,1]}(s_T) H(s_1, \ldots s_T). \tag{C.1}$$

$I_{[0,1]}(s)$ is the indicator function equal to 1 within the interval $[0, 1]$ and 0 outside of it; $H(s_1, \ldots, s_T) = 1$ if $s_1 \geq s_2 \geq \ldots \geq s_T$ and 0 otherwise. Representation (C.1) is convenient as it allows to calculate average values of order statistics by simple integration. Let us introduce the notation

$$\langle \phi(s_1, \ldots, s_T) \rangle = T! \int_0^1 ds_1 \int_0^{s_1} ds_2 \ldots \int_0^{s_{T-1}} ds_T \phi(s_1, \ldots, s_T),$$

where ϕ is an arbitrary function of the stimuli. Essentially, operation $\langle\cdot\rangle$ is nothing but taking the expectation value with the ordering $s_1 \geq s_2 \geq \ldots \geq s_T$. There are $T!$ possible orderings, and Eq. (C.1) arbitrarily fixes one of them, hence the multiplier $T!$ before the integral. We obtain

$$\begin{aligned}
&\mathrm{E}(u_{\max}) \equiv \langle s_1 \rangle = \frac{T}{T+1}, \quad \mathrm{Var}(u_{\max}) \equiv \langle s_1^2 \rangle - \langle s_1 \rangle^2 = \frac{T}{(T+1)^2(T+2)}, \\
&\mathrm{Cov}(u_{\max}, u_t) \equiv \langle s_1 s_t \rangle - \langle s_1 \rangle \langle s_t \rangle = \frac{T+1-t}{(T+1)^2(T+2)}, \\
&\mathrm{Cov}(u_{\max}, U_{\mathrm{A}}) = \mathrm{Cov}(u_{\max}, u_T) \equiv \frac{1}{T}\sum_{t=1}^{T} \mathrm{Cov}(u_{\max}, u_t) = \frac{1}{2(T+1)(T+2)},
\end{aligned}$$

and eventually the correlation coefficient defined by Eq. (5.3) is

$$r = \frac{\frac{1}{2(T+1)(T+2)} + \frac{1}{12T}}{\sqrt{\frac{1}{12T}\left(\frac{T}{(T+1)^2(T+2)} + \frac{1}{(T+1)(T+2)} + \frac{1}{12}\right)}}. \tag{C.2}$$

C.2 Anchoring-and-adjustment and adaptation processes

Consider a stream of stimuli s_t, $t = 1, \ldots, T$, that creates a temporally extended experience $(u_1, \ldots, u_T)$ T moments long. One form of anchoring-and-adjustment (Hogarth & Einhorn, 1992; Hands & Avons, 2001) describes the evolution of moment utilities as

$$u_t = u_{t-1} + (1-\alpha)(g(s_t) - u_{t-1}), \tag{C.3}$$

or,

$$u_t = \alpha u_{t-1} + (1-\alpha) g(s_t), \tag{C.4}$$

where $\alpha \in [0, 1]$, and $(1-\alpha)$ is the adjustment weight. As seen from Eq. (C.4), utility at moment t is a weighted average of utility at the previous moment and the current transformed stimulus. Formal solution of Eq. (C.4) gives

$$u_t = u_0\alpha^t + (1-\alpha)\sum_{k=0}^{t-1} \alpha^k g(s_{t-k}). \tag{C.5}$$

The formulation of the adaptation model that we use is due to Frederick and Loewenstein (1999). In this model the moment utility in period t is given by

$$u_t = g(s_t) - A_t, \tag{C.6}$$

where $\{A_t\}$ is a sequence of adaptation levels. Adaptation levels follow a process of anchoring-and-adjustment of the following kind:

$$A_t = (1-\beta)A_{t-1} + \beta g(s_{t-1}), \tag{C.7}$$

with initial condition $A_0 = 0$. Parameter $\beta \in [0, 1]$ determines the speed of adaptation. From Eqs. (C.6) and (C.7) it follows that

$$u_t = (1-\beta)u_{t-1} + g(s_t) - g(s_{t-1}), \quad t = 1, \ldots, T. \tag{C.8}$$

Here we set $g(s_0) = 0$. If $\beta = 0$, there is perfect adaptation ($A_t = 0$ for all t), and the evaluation simply reflects the current stimulus level: $u_t = g(s_t)$. If $\beta = 1$, then utility is only derived from a net gain in transformed stimulus, $u_t = g(s_t) - g(s_{t-1})$, $t = 1, \ldots, T$, and becomes negative if stimulus intensity drops.

Formal solution Eq. (C.8) has the form

$$u_t = u_0(1-\beta)^t + \sum_{k=0}^{t-1}(1-\beta)^k(g(s_{t-k}) - g(s_{t-k-1})), \quad t = 1, \ldots, T. \tag{C.9}$$

Author Note

Address: Department of Economics and Business, Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, 08005 Barcelona, Spain

E-mail: irina.cojuharenco@upf.edu

Tel: +34 93 542 2670

Fax: +34 93 542 2533