Can ValueFlow predict likeability of a commercial? An investigation of moment-to-moment affective reactions and the relation to marketing metrics

I.G.D. Dekker

09-07-2019

i.g.d.dekker@students.uu.nl

Applied Cognitive Psychology

Thesis Internship 27,5 ECTS

Validators

External Supervisor: Erik Prins, erik.prins@validators.nl

University Utrecht Supervisor: Martijn Mulder, m.j.mulder@uu.nl

Abstract

The purpose of this study is to investigate the integration of moment-to-moment affective experiences of a commercial into an overall evaluation. The company Validators has developed a moment-to-moment measurement tool, called ValueFlow, in which during the course of a commercial a person indicates their positive or negative affect to obtain insight into the dynamic structures of an television advert. The focus of the study is whether certain parts or key elements based on the psychological heuristic 'peak-end rule', have more influence in creating a well-liked commercial. Also, an explorative study was conducted to investigate to which extent the moment-to-moment affective experiences transfer to other well-known marketing metrics and how ValueFlow compares to overall commercial liking. Contradictory to existing literature, results of the study show no evidence for an effect of the peak-end rule. The moment-to-moment affective reactions show strong relationships with brand liking, call-to-action and branding. For the moment-to-moment affective experiences, a comparable transfer was found as overall ad liking, indicating further validation of the use of ValueFlow as a valid marketing measurement tool.

Keywords: Integration, moment-to-moment responses, overall ad liking, marketing, pre-testing

Table of Content

1. Introduction

1.1 Theoretical Background

Advertising researchers have been eager to find methods to measure the effectiveness of commercials. One way of testing the performance of a commercial is measuring to which extent a commercial is liked (Leather, McKechnie & Amirkhanian, 1994). In 1990 the Advertising Research Foundation (ARF) found that overall ad liking (OAL) was one of the marketing measurement tools which could best predict marketing success. Many studies have replicated the results of the ARF (Aaker & Stayman, 1990; Biel, 1990; Biel & Bridgewater, 1990, Dubitsky & Walker, 1994; Du Plessis, 1994a,1994b; Eagleson & Rossiter, 1994; Greene, 1992; Haley, 1990,1994, Haley & Baldinger, 1991, 2000; Smit, Van Meurs & Neijens, 2006).

However, since a commercial consists of different elements, insight into whether certain elements are more liked or disliked can be crucial for improving the design of a commercial. Reporting the overall likeability of a commercial provides no information about the internal dynamic structure of the viewer's experience of a commercial. Insight into the dynamic structures of a commercial can be obtained using a moment-to-moment measurement tool, in which a person indicates their positive or negative affect during a commercial (Aaker, Stayman & Hagerty, 1986). To measure the momentary affective responses, the company Validators has developed such a moment-to-moment measurement, called the ValueFlow. The ValueFlow is an online tool, in which a person indicates their current liking from moment-to-moment of a commercial by moving a slider at the lower part of the screen using a computer mouse. The data of the moment-to-moment responses of multiple participants is averaged and plotted over time. Thereby, allowing Validators to identify the parts of the commercial that are the least or most interesting to the viewers. Whereas a large amount of data is generated using this measure research showing which data predict a well-liked commercial and to which extent the ValueFlow is a reliable marketing measurement, is limited.

To investigate the above mentioned, the research has been split up into two parts. The first part focuses on which the data of the ValueFlow predicts a well-liked commercial. The second part of the study examines the extent to which the ValueFlow transfers to other well-known marketing pre-testing measurements and how ValueFlow compares to overall ad liking.

1.2 Part 1: Which data of the ValueFlow predicts a well-liked commercial?

The data of the momentary affective responses is fluctuating over time because of differences in likeability over different parts and specific moments. It is unclear which elements of the data predict an overall liked commercial. For the company Validators, it is important to know which patterns of affective response results in a well-rated and likeable commercial. Understanding which part of a commercial has the biggest effect on the overall liking of a commercial will provide important information as to possible improvements which can be made to the design of a commercial. To this end, this research will investigate whether some parts of a commercial have more influence on the overall likeability than others. Therefore, the beginning, middle and end of a commercial will be compared to understand which part has the most effect on its overall likeability. Subsequently, this research will investigate whether certain key moments, like the peak experience and the final moment in the commercial, have more weight in deriving a global evaluation.

1.2.1 From momentary experience to an overall evaluation

Studies on how a person gives an overall evaluation of past experiences differ in their outcomes. Some research shows that overall evaluations represent the average throughout the experience. Research of Aaker, Stayman and Hagerty (1986), Polsfuss and Hess (1991), and Thorson and Friestad (1989) suggests that a person determines a mean value of their responses over the commercial as a proxy measure for their appraisal of a commercial as a whole[1]. When a person estimates a mean value over their responses to make an evaluation for the overall appraisal of a commercial as Aaker et al. (1986), Polsfuss et al. (1991), and Thorson et al. (1989) suggest, then it is

---

[1] In this research, the appraisal- or overall evaluation of a seen commercial is called: 'Overall ad liking' (OAL).

expected that the 'beginning', 'middle' and 'end' of a commercial will equally well predict overall liking of the commercial.

However, Kahneman, Fredrickson, Schreiber and Redelmeier (1993) provide substantial support for peoples' use of the psychological heuristic called the "peak-end rule" for determining a judgement over an experience. They found that the evaluation of pain is influenced by the peak-end-rule, where the peak and the final levels of pain have the most influence in evaluating the experience. Based on the peak-end rule it is expected for this study that the 'end' will have more influence in creating OAL.

The peak-end rule occurs in many domains and with a diversity of different stimuli (Carmon & Kahneman, 1996; Redelmeier & Kahneman, 1996; Ross & Simonson, 1991; Loewenstein & Prelec, 1993; Loewenstein & Sicherman, 1991). However, only in one study the effects of the peak-end rule for commercials is examined (Baumgartner, Sajun & Padgett, 1997). Baumgartner et al. tested thirty commercials and found higher correlations for overall ad liking for the peak and final moment of the affective responses. This is in comparison with the sum of the affective responses. Their research supports the idea that the end of a commercial has a strong influence on the overall likeability. On the other hand, the beginning or middle of a commercial could also have a strong effect on the overall likeability, depending on whether the peak of the experience occurs in one of these parts. A side note to this research is that it was carried out at a University, with 30 students as participants, it is, therefore, unknown to which extent results could be generalized to the entire population, which should be the case for pre-testing commercials on a commercial scale.

In addition to the effect of the peak-end rule, the primacy and recency effect initially described by Murdock and Bennet (1962) could also play a role. The primacy and recency effect may result in the beginning and end of a commercial having a greater influence on the overall evaluation of a commercial than the middle part of the commercial because these parts would be better remembered.

In brief (Table 1), some research shows that the overall experience can be explained by the average of the experience. However, substantial evidence also points towards the peak-end rule, where the peak and final moment of the experience weigh more in the overall evaluation of that experience. However, the primacy and recency effect could also play a role in the determination on the evaluation of a commercial.

	Predictive-based on	Authors	Hypotheses
Beginning	Primacy effect	Murdock & Bennet (1962)	1
Middle	Peak occurring during this part	Kahneman et al. (1993) & Baumgartner et al. (1997)	1
End	Recency effect & peak-end rule	Kahneman et al. (1993) & Baumgartner et al. (1997) / Murdock & Bennet (1962)	1
Peak experience	peak-end rule	Kahneman et al. (1993) & Baumgartner et al. (1997)	2-a
Final Moment	peak-end rule	Kahneman et al. (1993) & Baumgartner et al. (1997)	2-b
Sum/average of all experiences	Mean value responses	Aaker et al. (1986), Polsfuss et al. (1991) & Thorson et al. (1989)	2

Table 1. Sum of theories and their authors about the predictive power of different parts and key elements for overall ad liking.

1.2.2 The predictive power of the 'beginning', 'middle' and 'end' of a commercial

The relation between the different parts to OAL will be investigated in determine whether one better predicts OAL than another. The end of a commercial is expected to be highly predictive based on the peak-end rule of Kahneman et al. (1993) and Baumgartner et al. (1997) and the primacy and recency effect of Murdock et al. (1962). The beginning of the commercial may be predictive for OAL because of the primacy effect. However, it is unknown whether the beginning or middle part of a commercial may also predict OAL, caused by the peak experiences. Controlling for in which part the peak will occur is difficult because it is unknown beforehand at what time during a commercial people experience their highest liking towards the commercial. This limitation, for the current study, of not having the option of controlling where the peak occurs makes it difficult to draw a definite conclusion on the role of the peak for which part of the commercial is most predictive. However, with the

combination of the recency effect and the peak-end rule, it is expected that the end of the commercial would be more predictive than the beginning and middle part of a commercial. The following hypothesis is drawn upon.

H_1 : *Overall Ad Liking will have a higher correlation with the end of a commercial than the beginning and middle part of a commercial.*

1.2.3 The peak, final moment & average over the experience (ValueRate)

Further analyses of the peak experience and final moment independently of the beginning, middle or end, can provide further knowledge about the role of these variables. Kahneman et al. (1993) suggest that a person does not sum up the experiences, but relies on a weighted average model where certain key moments, like the peak and the final moment weigh more in the model for prediction of OAL. This model of Kahneman was tested for commercials by Baumgartner et al. (1997). They compared the peak and final moment with the sum of the momentary affective experiences (SMAE) and found, as hypothesised, higher correlations for those key elements. However, Baumgartner did not apply a correction for the length of a commercial. Not correcting for the length means the amount of data increases alongside the duration of a commercial and therefore a higher sum of affective experience. Since different sizes of commercials are used in this study a correction for the length of the commercial has been applied by dividing the sum of the momentary affective experience with the length of the commercial creating a new variable called ValueRate. All in all, the ValueRate will be used in the current study as a comparison baseline to investigate whether the peak experience and final moment are indeed more predictive key elements. Even though Baumgartner should have applied a correction for the duration of a commercial, it is still expected, also based on the model of Kahneman (1993) and other research on the peak-end rule, that these key moments should correlate higher with OAL than the ValueRate. The following hypotheses have been drawn upon.

H_{2-A} : *Overall Ad liking will significantly be higher correlated with the peak experience than with the ValueRate.*

H_{2-B} : *Overall Ad liking will significantly be higher correlated with the final moment than with the*

ValueRate.

Additionally, for the company Validators, an exploratory analysis was conducted. For this analysis, a multiple regression model, based on the idea of Kahneman et al. (1993), was used. The assumption being that people rely on a weighted average model wherein certain key elements are weighted more heavily than others. The idea is that a person does not rely on one source of information, but uses multiple weighted sources of information and integrates them to come to one outcome. This same principle could also work for the data of the ValueFlow. Creating a model consisting of multiple predictors may better predict OAL than the predictors alone. This will be tested by creating a regression model with a combination of the ValueRate, peak and final moment. It is expected that the model would make a better prediction for OAL than the individual factors alone.

The main goal of part one is to determine which data of the ValueFlow predict a generally well-liked commercial. Additionally, it will give insight into understanding whether certain parts or key elements are more influential in the integration of moment-to-moment experiences into a global commercial judgment.

1.3 Part 2: Transfer to other marketing measurements

The second part of this study consists of an exploratory study for the company Validators. The goal is to provide more insight into the extent to which the ValueFlow transfers to other well-known marketing pre-testing measurements and how the ValueFlow compares to overall ad liking. For a measurement tool generally, it is important to have to some degree strong relationship with other marketing measurement tools. Because such tools all try to predict, in their own way, the same concept, namely marketing success (Leeflang & Wittink, 2000). Therefore, the expectation is that the ValueFlow will correlate positively with other marketing measurements. Baumgartner et al. (1997) already studied 'brand liking' and 'brand recall' and found significant correlations between these measurements. On the other hand, ad likeability is also found to have a poor correlation with brand recall (Smit, Van Meurs & Neijens, 2006). The current study expands the pre-testing

measurements with 'branding', 'call-to-action - purchase intentions' and 'call-to-action - intention to get more information'. For commercial 'branding', how strong the brand is linked to a commercial is measured (Young, 2008). Young (2008) states that a commercial should avoid being universal. A poorly designed commercial is one in which any similar brand could be used and it would not make a difference. An adequately branded commercial is when a person could not describe the commercial without contemplating about the brand. It is expected that a commercial which scores higher on likeability will also score higher on branding. Call-to-action is also a popular way to measure persuasion, which is a factor strongly related to sales (Haley & Baldinger, 1991). A good commercial creates a sense of urgency that motivates the call to action (Young, 2008). Shim, Eastlick, Lotz and Warrington (2001) described that especially measuring the intention to act in the form of purchase is a strong predictor for commercial marketing success. Kleins (1998) mentions the importance for companies to measure the intention of the consumer to search for more information after seeing the commercial. For both call-to-actions, it's expected that a more liked commercial will lead to higher intentions to act.

In summary for part two it is important to know, for the company Validators, whether the ValueFlow transfers equally well or even better to other marketing metrics as asking overall ad liking (OAL). Reminding that OAL is one of the strongest factors predicting marketing success (Haley and Baldinger, 1991; Smit, Meurs & Neijens, 2006) and the ValueFlow does basically measure the same concept, but on a momentary level a comparable transfer of the ValueFlow towards other marketing measurement tools would further validate the use of a ValueFlow as a valid marketing measurement tool.

2. Methodology

2.1 Subjects & Procedure & Stimuli

600 participants were included in the experiment. Their age ranged from 18 to 65 years (M = 43.98, SD = 13.14). Fifty-three per cent of the participants were female, forty-seven per cent was male. The participants were directed to the study via a panel service organization called '*Panel Inzicht*'. The experiment was conducted through a website on a desktop and sound was a requirement. Participants were informed that they had to indicate their liking towards five commercials and that questions about these commercial will be asked. Instructions how to value their liking towards the commercial using the ValueFlow tool was explained using an instruction video. Participants practised with the ValueFlow with one commercial. Between each commercial participants had to indicate their overall liking of it, brand liking, call to action and familiarity. After the commercials, the participants were given an unguided brand recall test.

Thirty advertisements were used in this study, which bears a sample of reasonable size. However, because showing thirty commercials would be too tiring for the participant, five out of thirty commercials were randomly selected and shown to each participant. In doing this every commercial was evaluated with the ValueFlow around a hundred times. Based on the experience of the company this is a reasonable amount to obtain valid data resembling the moment-to-moment liking of the Dutch people towards that commercial. The order in which the commercials were shown was randomized.

2.2 Apparatus and Measurement of Moment-to-Moment responses

Moment-to-moment affective responses were measured with a digital tool which the company Validators calls the ValueFlow. A commercial is displayed on the computer screen with a height of 366px and a width of 650px. On the right side next to the commercial, there is a vertical slider with anchors ranging from 'very positive' to 'neutral' to 'very negative'. The slider had a height of 366px and a width of 150px. A participant started the commercial by clicking a play button in the middle of

the video player. During the commercial, the slider did not have to be pressed and dragged, but by moving the mouse up and down the momentary liking towards the commercial could be given. Data was recorded when a movement of the mouse took place. A conversion over all the data of a commercial was done so every 100 milliseconds resulted in a data point ranging from -5 till 5 on a scale of likeability towards that moment of the commercial.

2.3 Dependent measures

The dependent variables range were measured with a web-based graphic-slider ranging from 0 to 100 (Cook, Heath, Thompson, 2001). Overall liking of the commercial was measured with the slider ranging from *"I don't like the commercial at all"* (0) to *"I like the commercial very much"* (100) brand liking was measured with a slider ranging from *"Very negative"* (0) to *"Very positive"* (100) (Cook et al., 2001). Call-to-action was broken down into two factors. One factor of call-to-action corresponded to taking action in buying the product of a brand. The following question was asked: *'Did the TV commercial encouraged you to buy the products of 'brand X'?'* Possible responses ranged from *"Certainly Not"* (0) to *"Very Certainly"* (100) (Cook et al., 2001). The other factor of call-to-action resembled taking action in the form of searching information about 'brand X' and the following question was asked: *'Did the commercial encouraged you to search more information about 'Brand X'?'* The slider ranged from *"Certainly Not"* (0) to *"Very Certainly"* (100) (Cook et al., 2001). A check for interest in the product was done by asking whether a person is interested in the products of Brand X regardless of the commercial, with a slider ranging from *"Certainly Not interested"* (0) to *Very Interested (100).* For measuring commercial branding the following question was asked: *"To what extent does the commercial displays that it is a commercial for brand X"*. With a slider ranging from "*This commercial could be for a variety of brands*" (0) to "*You couldn't help but think this commercial is for 'brand X'"* (100). For checking for familiarity participants were asked to indicate whether they had seen the commercial on a three-point Likert scale ranging from *"Certainly not"* (-1) to *"Yes, maybe"*(0) to *"Yes, certainly"* (1). At the end of the

survey, the participants had to freely recall the brands they had seen, the order of which did not matter.

2.4 Analyses & Confound factors

For the check for familiarity, an average across subjects' scores of the 3 points Likert scale was taken per commercial. The computed familiarity score indicated whether more people had seen the commercial or not, ranging from -1 to 1. For example, a score towards -1 would indicate that more people had not seen the commercial and a score towards +1 would indicate that more people had seen the commercial. A score of 0 would indicate that just as many persons had seen the commercial as had not seen the commercial. An interaction analysis was conducted to check whether familiarity has influenced the data of the ValueFlow for predicting OAL.

Baumgartner et al. (1997) used the sum of the momentary affective experience as a baseline for comparison with the variables peak experience and final experience. However, taking the sum of all experiences would mean that a longer commercial, would have more momentary responses and therefore lead to a proportionately higher SMAE than a shorter commercial. Since different sizes of commercials have been used in this study a correction for the length of the commercial has been applied for SMAE. This correction was made by taking the area under the curve of the ValueFlow (Figure 1) and by dividing it with the length of the commercial creating a variable called ValueRate. An analysis has been conducted to show that the SMAE is affected by the duration, and ValueRate is not, by correlating these variables with the length of the commercial. The expectation being that the SMAE would have a significant relationship with the length of the commercial and ValueRate would not.
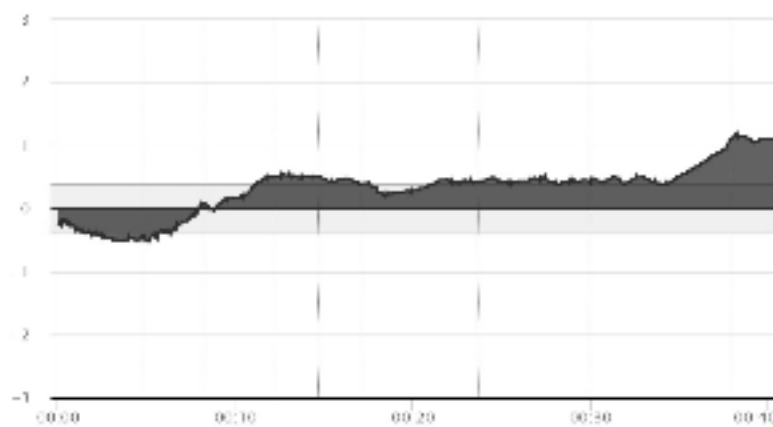
Figure 1. Example of the ValueFlow of a commercial. On the y-axis the moment-to-moment liking towards the commercial (Range -3 to +3). On the x-axis time (seconds). The grey area represents the area under the curve.

Another factor which may is creating a bias to the model is the first few seconds of the measurements. It may take a short-term for a person to form an impression about the commercial and to take a position with the mouse on the range from disliking or liking the commercial. In addition, a person always starts around zero and thereby the first few seconds of the data may be distorted. Not taking this into account and taking the data as if the data represent a valid momentary liking would mean that every first few seconds of a commercial would neither be liked nor disliked. While the peak and the final moment are not affected by the first few moments, the ValueRate is. Therefore, not including the first few moments would result in a significantly better model to predict OAL. The following approach was used: the time until the first turning point in the data was taken (figure 2), because it seemed that this was the point where a stable opinion about the commercial was formed. Because the time until the first turning point differs in commercials, the average duration, of 3.74 seconds (SD=1.55), until the first turning point was taken. An analysis was conducted to show that for ValueRate the beginning does indeed creates bias to the model for prediction of OAL. Therefore, a comparison was made for the data with and without the first few moments. This was to determine whether indeed the data of the ValueFlow without the noisy beginning significantly better predict the OAL.
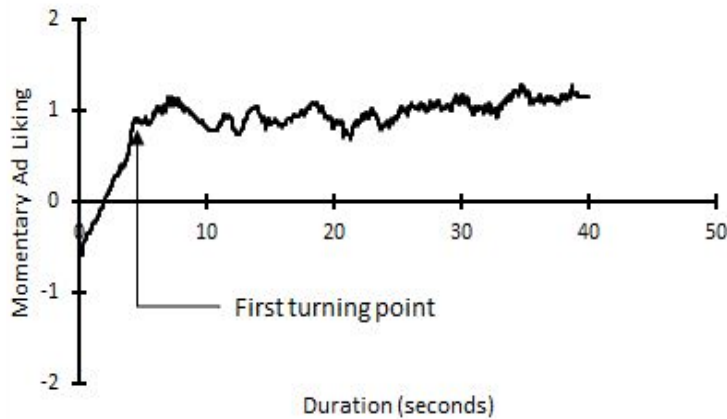
Figure 2. Example of the first turning point of the response trace of a commercial. The turning point seems to indicate that a stable opinion was formed. On the y-axis the moment-to-moment liking towards the commercial (Range -3 to +3). On the x-axis time (seconds).

3. Results

Based on the previous work of Van den Abeele and MacLachlan (1994) and Baumgartner et al. (1997), the commercials were selected as the unit of analysis. Therefore, an average over the responses was taken. Forty participants were removed from the analysis[2]. Leaving 560 participants for the analysis.

3.1 Descriptive findings

Figure 3 displays the average moment-to-moment affective appreciation of the participants of the 30 commercials. The graphs show that for most of the commercials the participants had a general positive moment-to-moment attitude towards the commercials (N = 26). Though, four commercials show a negative tendency. Noticeable is that the commercials show fairly flat lines, with a maximum peak of 2,46 (Range -5 to 5). The average maximum peak over the 30 commercials was 1,20 (SD = 0,59).

---

[2] Non serious answers were given to an question in which participants were asked to describe their favourite scene. These data were used for another study for the company Validators.
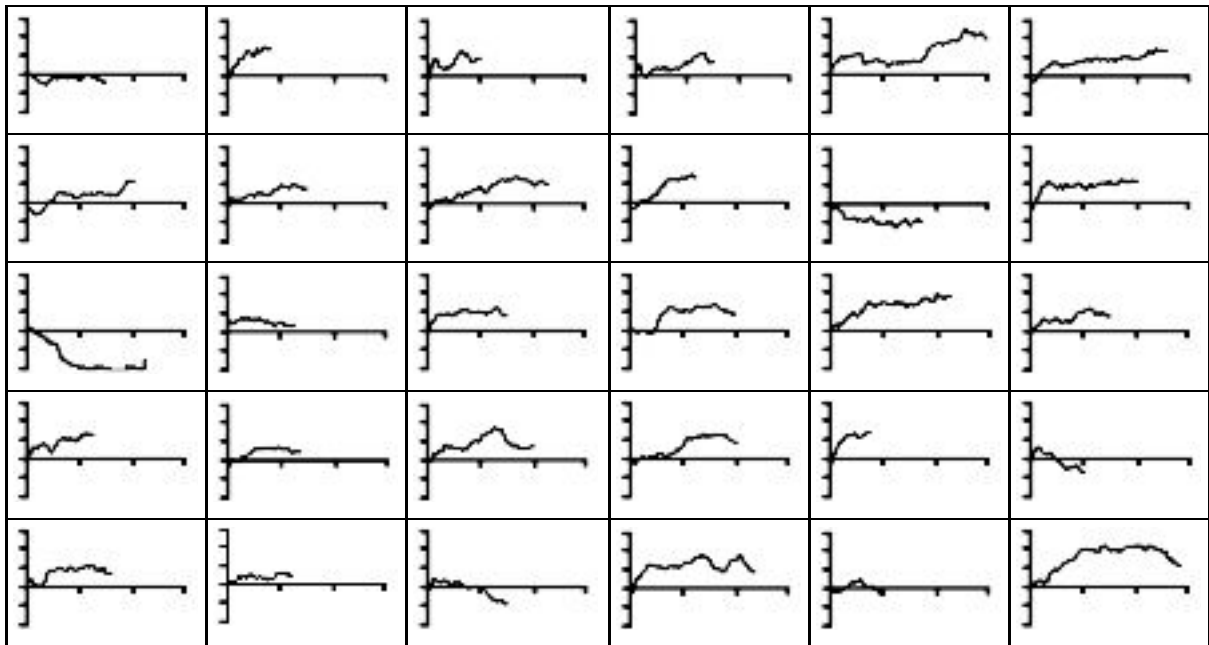
Figure 3. Affective trace of the commercials. Horizontal axis length of the commercial in seconds (0 to 60); vertical axis momentary affective likeability of the commercial (range -2 to 3).

3.2 Part 1: Which data of the ValueFlow predicts a well-liked commercial?

3.2.1 The predictive power of the 'beginning', 'middle' and 'end' of a commercial

To test which part of the commercial is the most predictive for overall ad liking the average over the area under the curve of the ValueFlow of these parts was taken and compared to determine which part is the most predictive for overall ad liking. Without a standardized rule for defining what the beginning, middle or end of a commercial is, the decision was made to divide those parts equally in fractions of ⅓. The peak occurred in 83% of the commercials during the ending part of the commercial. In 13% of the cases, the peak occurred in the beginning. Striking is that for these cases the ValueFlow showed a negative trend line(e.g. Figure 4). Only one commercial had the peak in the middle.
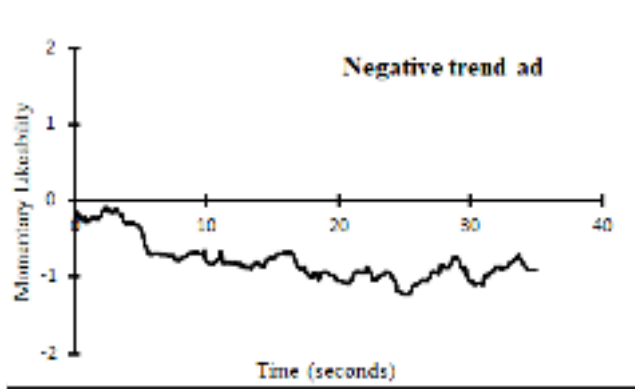
Figure 4: Example of a commercial with a negative trend. Horizontal axis length of the commercial in seconds (0 to 40); vertical axis momentary likeability of the commercial (range -2 to 2).

The zero-order correlations between the different parts and the dependent variable of overall ad liking are shown in Table 2. All parts show a significant relationship with OAL $r_s > .462$, 95% BCa CI [-.053, .9967], $p < .01$ and also after control for familiarity the results show a significant result with $r_s > .477$, 95% BCa CI [-.024, .971], $p < .01$. T-statistic was computed to compare whether there is a significant difference between the correlations of the different parts based on the method of Chen & Popovich (2002). The results reject the first hypothesis H_1, because no significant difference was found for the correlations with OAL of the middle part and the end, difference = -.063, $t(27) = -1.352$, $p = .09$. Yet, a significant difference was found for the correlation between the begin and OAL and mid and OAL, difference = -.259, $t(27) = -2.960$, $p = .003$. Also, the correlation between the begin and OAL was significant different from the correlation between the end and OAL difference = -.322, $t(27) = -3.717$, $p < .001$.

	Begin		Mid		End	
	Zero-order Correlation	Partial Correlation	Zero-order Correlation	Partial Correlation	Zero-order Correlation	Partial Correlation
Likeability N = 30	.462* [-.053, .758]	.477** [-.024, .759]	.721** [.148, .934]	.718** [.187, .940]	.784** [.259, .967]	.796** [.389, .971]

Table 2. Pearson's r correlations between different parts of the commercial (begin, mid & end) and overall likeability score. Partial correlation controlling for the variable familiarity to get a pure measurement of the relationship between the different parts of the commercial and variability. Ns = not significant (p > .05), * = p <0,05, ** = p <0.01,. BCa bootstrap 95% CLs reported in brackets.

3.2.2 The peak, final moment & average over the experience (ValueRate)

The following predicting variables were computed, based on the research of Baumgartner et al. (1997) for testing the hypotheses H_{2-A} and H_{2-B}: the peak experience (i.e., the highest moment-to-moment likeability trace); final moment (i.e the height of the trace during the last moment of the commercial); the sum of the momentary affective experiences (i.e., the sum of all the deviations from neutral of the trace; SMAE); ValueRate (i.e. area under the curve divided by the duration of the commercial). Pearson correlation coefficients were computed (Table 3). All factors (peak experience, final moment, SMAE, Value Rate) were significantly related to the OAL r_s > .767, 95% BCa CI [>.413, <.970], p < .001 even after controlling for familiarity results show a strong relationship with the OAL r_s > .770, 95% BCa CI [>.440, <.978], p < .001. Controlling for familiarity shows that the results are robust and are not altered by the involvement of familiarity. T-statistic was computed to compare whether there is a significant difference between the Pearson correlations calculated with the method of Chen & Popovich (2002) (Table 4). The results reject both hypotheses $H_{2-A \& B}$. The correlation between the ValueRate and OAL was significantly higher than the correlation between the peak and OAL, difference -,09 $t(27)$ = -1,812, p = .04. The correlation between the final moment and

OAL was not significantly lower than the correlation between the ValueRate and OAL, difference = -0.023 $t(27) = -.605$, $p = .274$. See Table 4 for the p-values of the difference between the other factors.

	Peak		Final moment		SMAE		ValueRate	
	Zero-or der C.	*Partial C.*	*Zero-ord er C.*	*Partial C.*	*Zero-ord er C.*	*Partial C.*	*Zero-or der C.*	*Partial C.*
Overall ad liking	.767 ** [.415, .913]	.770 ** [.471, .922]	.834 ** [.415, .952]	.841 ** [.535, .954]	.811 ** [.413, .946]	.850 ** [.444, .944]	.858*** [.488, .970]	.856*** [.541, .968]

Table 3. Pearson's r correlations between different predictors variables (Peak, Final Moment, Sum of Momentary affective experience (SMAE) & ValueRate) and overall likeability score. Partial pearson correlation controlling for the variable familiarity to get a pure measurement of the relationship between the different parts of the commercial and variability. Ns = not significant p > .05, * p <0,05, ** p <0.01 BCa bootstrap 95% CLs reported in brackets.

	Peak	Final moment	SMAE	ValueRate
Peak	x	.77	.208	.040
Final moment	df = 27	x	.657	.274
SMAE			x	.125
ValueRate				x

Table 4. T-statistic with significance p-values comparison Pearson correlations between different predicting variables (Peak, Final moment, Sum of momentary affective experience (SMAE) & ValueRate) calculated with the method of Chen & Popovich (2002).

3.2.3 Regression model

To control whether a regression model makes a better prediction than the predicting variables alone. The following predictors were included in the model: ValueRate, the peak and the final moment. Results show high collinearity with an average VIF of 7.509 and average tolerance of .138. Besides the collinearity diagnostics, the peak and end did not significantly contribute to the model with both predictors showing the significance of $p > 0,28$.

	b	SE B	β	t	p	tolerance	VIF
Constant	50.07 [46.45 , 54.14]	3.13		15.95	p < 0.001		
ValueRate	9,636 [.794 , 14.69]	4.19	.607	2.30	p = 0.030	0.140	7.153
Peak	-1,047 [-11.61 , 13.41]	3.89	-.065	-0.27	p = 0,79	0.169	5.929
End	3,642 [-1.61 , 7.458]	3.32	.332	1.10	p = 0,28	0.106	9.447

Tabel 5. A linear model of predictors of overall ad liking, with 95% bias-corrected and accelerated confidence intervals reported in parentheses. Confidence intervals and standard errors based on 1000 bootstrap samples.

3.2.4 Familiarity check

No significant results were found for the interaction between familiarity and the data of the ValueFlow. The results are displayed only for the interaction between ValueRate and familiarity in Table 5 and Figure 5, the other variables show similar results and no effect was found for the other predictor variables: Peak * Familiarity b = .72, SE B = 5.9709 , *t* =.12, p =.904; Final moment * Familiarity b = .72, SE B = 5.9709 , *t* = .12, p =.904.

	b	SE B	t	p
Constant	56.74 [54.37, 59.10]	1.15	49.28	p < 0.001
Familiarity	2.2117 [-6.24, 10.66]	4.11	.53	p = .595
ValueRate	13.3790 [9.49, 17.26]	1.89	7.07	p < .001
Familiarity x ValueRate	—.8082 [-11.73, 10.12]	5.31	—.15	p =.880

Table 6. R^2 = .74 Linear regression and interaction effect of familiarity and ValueRate.
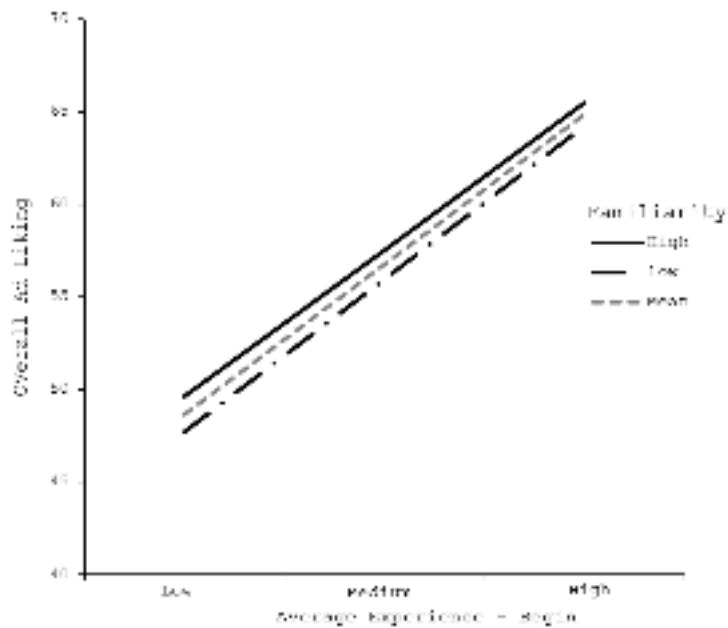
Figure 5. Simple slopes equations of the regression of ValueRate on overall ad liking at three levels of familiarity. No interaction effect visible.

3.2.5 Confound factors: ad duration & beginning of the measurement

Results of the analyses show that SMAE is affected by the duration of a commercial with a significant correlation between these variables $r_s = .511$, 95% BCa CI [.038, .865], $p < .0$. ValueRate shows no significant relationship with duration of a commercial $r_s > .236$, 95% BCa CI [-.135, .679], $p < .01$.

The results of the analyses show that for both the SMAE and the ValueRate taking the first 3.74 seconds of the commercial leads to a significant improvement to the relationship between these variables and OAL. The correlation between the ValueRate with the first 3.74 seconds and OAL was significantly lower than the correlation between the ValueRate without the first 3.74 seconds and OAL difference = -0.0075 $t(27) = -1.769$, $p = .044$.

3.3 Part 2: Transfer to other marketing measurements

Data was collected for multiple well-known marketing measurements, consisting of brand liking, brand recall, CTA — purchase intention, CTA — intention to search for more information and branding. The results are presented in table 7. All the different proxy predictors, except brand recall,

show strong and significant zero-order correlation with the different marketing measurements $r_s >$ ..368, $p < .05$. Also when correcting for familiarity a significant relationship is found $r_s > .368$, $p < .05$. Controlling for familiarity shows that the results are robust and are not altered by the involvement of familiarity. The different proxy predictors show no significant relationship with brand recall $r_s < .077$, $p > .05$. After controlling for product interest the proxy predictors still show a significant relationship with both CTA's $r_s > .350$, $p < .05$, except for the relationship between the peak experience and CTA - purchase intention $r_s < .077$, $p > .05$. The predictor 'final moment' showed the highest correlation with the marketing measurements, however, this difference was not significantly different from the correlation between the ValueRate and the marketing measurements, with a maximum difference of 0.096, $t(27) < -.781$, $p > .05$

	Brand liking		Brand recall		CTA- purchase intention			CTA — intention search information			Branding	
	Zero-order C.	Partial C.	Zero-order C.	Partial C.	Zero-order C.	Partial C.	Partial C. Inter.	Zero-order C.	Partial C.	Partial C. Inter.	Zero-order C.	Partial C.
OAL	.795 ** [.652, .877]	.788 ** [.650, .873]	.077 Ns [-.295, .431]	.025 Ns [-.347, 413]	.695 ** [.397, .862]	.684 ** [.397, .870]	.613 ** [.047, 833]	.794 ** [.558, .905]	.803 ** [.540, .929]	.723 ** [.333, .880]	.580 ** [.052, .832]	.560 ** [.062, .810]
Value Rate	.636 ** [.304, .826]	.630 ** [.320, .822]	-.095 Ns [-.414, .224]	-.137 Ns [-433, .164]	.629 * [.293, .814]	.622 * [.291, .825]	.688 ** [.132, .870]	.718 ** [.415, .853]	.720 ** [.409, .868]	.701 ** [.268, .867]	.520 ** [-.095, .834]	.513 ** [-.60, .819]
Peak	.572 ** [.277, .780]	.571 ** [.299, .776]	.035 Ns [-.333, .400]	.017 Ns [-.341, .367]	.460 * [.102, .757]	.457 * [.117, .755]	.350 Ns [-.195, .672]	.528 ** [.187, .738]	.575 ** [.156, .764]	.431 * [.017, .711]	.368 * [-.082, .707]	.368 * [-.055, .694]
Final Moment	.732 ** [.417, .901]	.738 ** [.469, .897]	-.165 Ns [-.505, .187]	-.185 Ns [-.522, .194]	.688 ** [.365, .895]	.693 ** [.393, .898]	.545 ** [-.023, .825]	.751 ** [.443, .893]	.751 ** [.421, .909]	.629 ** [.194, .836]	.564 ** [.049, .845]	.582 * [.106, .842

Table 7. Pearson's r correlations between overall ad liking (OAL), different predictor variables (ValueRate, peak experience & final Moment) and different marketing measurements, (brand liking, brand recall, call-to-action - purchase intention, call-to-action - intention to search information & branding. Partial correlation controlling for the variable familiarity to obtain a pure measurement of the relationship between predicting variables and marketing measurements. For call to action partial controlling has also been applied for interest in product regardless of the commercial. Ns = not significant p > .05, * p <0,05, ** p <0.01. BCa bootstrap 95% CLs reported in brackets.

T-statistic was computed to investigate whether there is a significant difference between the Pearson correlations of the ValueFlow proxy predictors and Overall ad Liking calculated with the method of Chen & Popovich (2002). The results (Table 8) show that the ValueRate and final moment do evenly well predict the different marketing measurements as overall ad liking does, with a maximum difference of .159 $t(27) = -1.769$, $p > .123$, except for brand recall. The correlations of brand recall to the marketing measurements were not significant, so the results of the comparison whether the ValueFlow transfers, in the same way, are not included in the analyses.

	Brand liking	CTA- purchase	CTA - information	Branding
Value- Rate	.993	.193	.123	.241
Peak	**.004**	**.009**	**.001**	**.027**
Final Moment	.173	.434	.257	.429

Table 8. P-values of t-statistic for the comparison between correlations for ValueFlow Proxy predictors and overall ad liking to the different marketing metrics (brand liking, call-to-action purchase intentions, call-to-action intention to search for information & branding)

For the marketing measurements brand liking, branding, call-to-action purchase intention and intention to search more information overall ad liking, the final moment was the strongest predictor. These predictors were, however, not significantly different from the relationship between overall ad liking and these marketing measurements with a maximum difference of -0.063, $t(27) < -.956$, $p > .173$.

3.3.1 Product Interest Check

No significant results were found for the interaction between product interest and the data of the ValueFlow. For the interaction effect the variable, which had the highest relationship with call-to-action - purchase, has been chosen to be displayed (Table 9, Figure 6). However, similar results with no interaction effects were found for the other ValueFlow proxy variables. : ValueRate *

Interest b = -.085, SE B = .205, t =-.415, p =.681; Peak * Interest b = -.01, SE B = .2581, t = -.05, p =.957.

	b	SE B	t	p
Constant	43.19 [54.37, 59.10]	.79	49.28	$p < 0.001$
Interest	.701 [.46, .93]	.11	.53	$p = .595$
Final moment	2.66 [-1.26, 6.573]	1.91	7.07	$p < .001$
Interest x Final Moment	-.030 [-.21, .15]	.086	-.15	$p = .880$

Table 9. R^2 = .83 Interaction effect of Interest and Final moment on call to action - intention to search for more information.
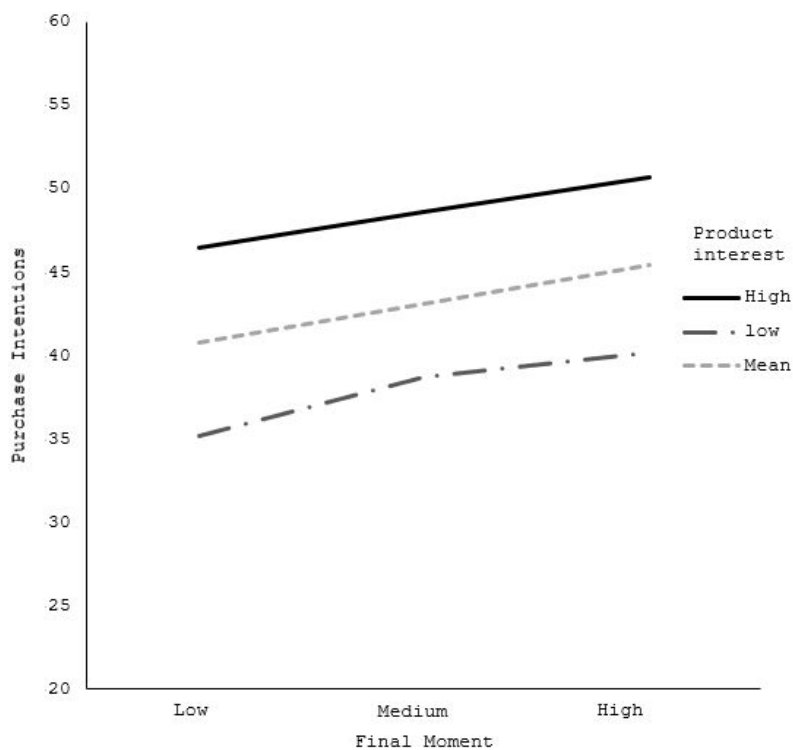

Figure 6. Simple slopes equations of the regression of the Final moment of VF on purchase intentions at three levels of Product interest. No interaction effect visible.

For the interaction of product interest with intention to search for more information. For the interaction effect the variable, which had the highest relationship with call-to-action - purchase, has been chosen to be displayed (Table 10, Figure 7). However, similar results with no interaction effects

were found for the other ValueFlow proxy variables : ValueRate * Interest b = -.09, SE B = .206, t =-.467, p =.64; Peak * Interest b = -.11, SE B = .2005, t = -.53, p =.598.

	b	SE B	t	p
Constant	39.77 [37.88, 41.67]	.92	43.12	p < 0.001
Interest	.36 [.13, .60]	.11	3.13	p = .004
Final moment	3.61 [.14, 7.09]	1.69	2.137	p = .04
Interest x Final Moment	-.04 [-.27, .19]	.112	-.39	p =.69

Table 10. R^2 = .83 Linear regression and interaction effect of Interest and Final moment on call to action -intention to search more information.
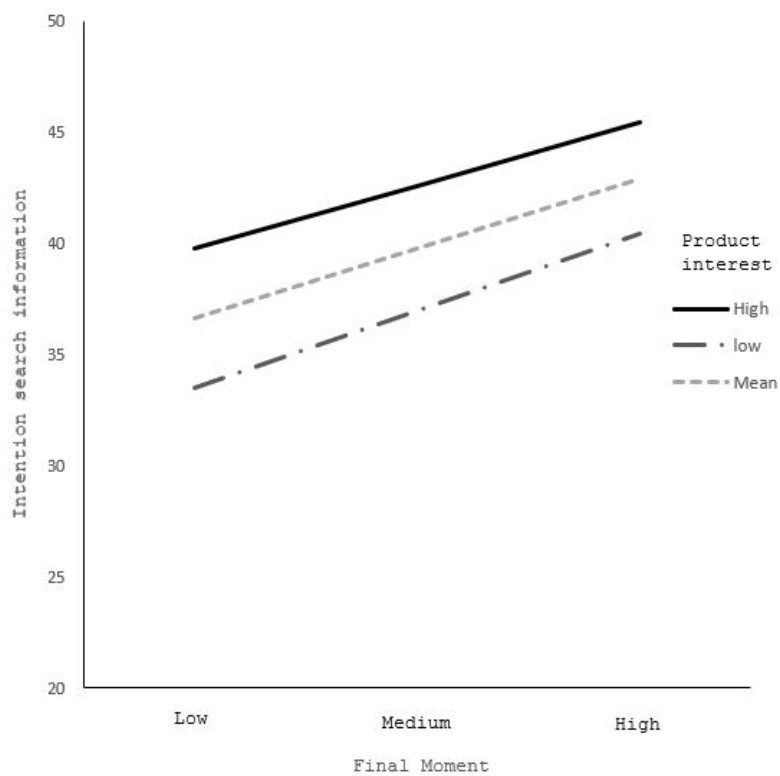

Figure 7. Simple slopes equations of the regression of the Final moment of VF on intentions to search more information at three levels of Product interest. No interaction effect visible.

4. Discussion

The present study has tried to investigate the integration of moment-to-moment affective experiences into an overall ad evaluation and to explore the transferability of these moment-to-moment affective experiences to other marketing measurements. In the first part of this research, the aim is on which data of the ValueFlow predict a commercial which is well-liked. Therefore, the beginning, middle and end of a commercial are compared to determine which part has the most effect on the overall likeability. Subsequently, whether certain key moments would have more weight in deriving a global evaluation than taking the mean over all the experiences was investigated. In the second part of the study, the extent to which the ValueFlow tranfers to other well-known pre-testing marketing measurements is analyzed. Subsequently, a comparison with the transfer to advertising likeability was made (OAL).

4.1 Part 1: Which data of the ValueFlow predicts a well-liked commercial?

4.1.1 The predictive power of the 'beginning', 'middle' and 'end' of a commercial

The results show that the 'end' of the commercial did not have a stronger relation to OAL than the 'middle' of a commercial. Therefore, rejecting H_1 which states that Overall Ad Liking (OAL) has a higher correlation with the end of a commercial than the beginning and middle part. This finding contradicts the expectations based on the research of Kahneman et al. (1993) and Baumgartner et al. (1997). Their results showed that the end of a commercial would be the most predictive based on the peak-end rule. However, the finding that the middle parts equally predict the OAL, though the peak did not occur in the middle section (except for one case), points out that the influence of the peak for the predictivity of the different parts in a commercial is not so solid as suggested by Baumgartner et al. (1997). The results of the current study suggest that the middle and end of a commercial are the same in predictability for OAL.

The results of this study also show a less predictive beginning in comparison to the rest of the commercial. This aspect can be explained by the noisy data of the first few moments of the measurement. However, it cannot be concluded that the first part of a commercial is less important for creating a likeable commercial. It seems that the beginning is also important for creating a likeable ad, but the ValueFlow method cannot validly measure the first few moments of a commercial and therefore resulting in less predictive data for predicting OAL. Further research is needed to determine to which it is possible to shorten the noisy data.

The primacy and recency effect initially described by Murdock and Bennet (1962) may result in the beginning and end of a commercial having a larger influence on the overall evaluation of a commercial than the middle part of the commercial. The reason being that the beginning and the end of a commercial are remembered better and therefore weigh more. Though the finding that the middle and the end part of a commercial equally predict the OAL does indicate that the primacy-recency effect did not play a significant role in the studied commercials.

4.1.2 The peak, final moment & average over the experience (ValueRate)

The results show that both the peak and the final moment did not have a significantly better relation to OAL than the ValueRate, meaning that for the data of the ValueFlow the peak or final moment is not a better proxy predictor for overall ad evaluation than the ValueRate. Therefore, both hypotheses H_{2-A} and H_{2-B}, which state that overall ad liking correlates significantly higher with the peak experience and final moment than with the ValueRate. These findings contradict the idea that in deriving an overall liking of a commercial, out of the data of the ValueFlow, the peak and the final moment would have a higher influence than the average over the responses. However, it is better to take the ValueRate as a proxy predictor to measure how a commercial will score on OAL. These results, however, may not reject the model of the peak-end rule of Kahneman and Colleagues (1993). On an individual level, it is still very plausible that a consumer focuses on a few key moments for deriving a global evaluation of a commercial. But, for this study, as in the study of Baumgartner et al (1997), where the commercials were taken as the unit of analysis and participants were used as

replicates by taking the average over their responses, no evidence was found for the theory that for predicting AOL the peak and the final moment of the data of the ValueFlow are more predictive.

The results of the current study differ from those found by Baumgartner et al. (1997). A possible explanation could be the difference in studied participants. In the current study participants with wide-ranging ages were included, so that the participants represented a national representative distribution. These participants conducted the experiment via a survey and Baumgartner et al. (1997) conducted his research among university students at the university. This difference could have had strong implications for the results and especially to the validity of taking an average over the participants and therefore using the commercials as a unit of analysis. In the paper of Baumgartner et al. (1997) the following requirement was stated for treating the participants as replicates: "Averaging across subjects requires that there be reasonable consistency in different subjects' responses to the advertisements in terms of both their moment-to-moment responses and overall judgments. The responses of 1 of the 28 participants were idiosyncratic, and this person was excluded from further analyses. The remaining 27 participants' responses were aggregated as they demonstrated high consistency on both the moment-to-moment responses across advertisements and the various dependent measures" (Baumgartner et al., 1997, p. 223). By using only students as participants Baumgartner et al. (1997) probably reached much higher consistency over the responses, than when a diverse test audience would have been, as was the case in the current study. This can be explained by the fact that students most likely like advertisements in a more similar way, than, for example, a young woman in comparison to an old man would. Whereas Baumgartner found high consistency in the subjects' responses and excluded a person to reach higher consistency, the responses in this study show no consistency (Figure 8).
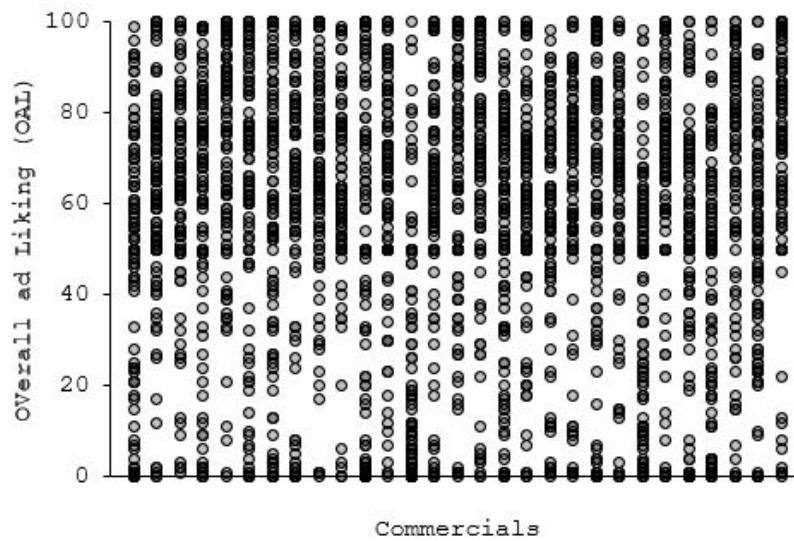
Figure 8. Consistency in responses of overall ad liking of the tested commercials. On the y-axis overall ad liking (0-100). On the x-axis the different commercials.

How can it be explained that the high consistency in responses found by Baumgartner et al. (1997) provide proof of an effect of the peak-end rule? The peak-end rule has been confirmed in many studies, but always on an individual level. Therefore, taking an average over the participants may only be justified, when the peak and final moment of the average trace actually represent the peaks and final moments of all the individuals, which may only be possible when high consistency is present in the responses of the participants, which was the case in the study of Baumgartner et al. (1997). However, since this study shows very low consistency in responses and the location of the peak of the individuals differed substantially from the location of the peak of the average trace (Figure 9). The conclusion can be drawn, that not finding proof of an effect of the peak-end rule in this study could be explained by the fact that the peak and the final moment of the average trace did not represent the actual peaks and final moments of the individual participants. Further research is needed to determine the extent to which a difference between the individual response and the average response for the peak and final moment influences the peak-end rule.
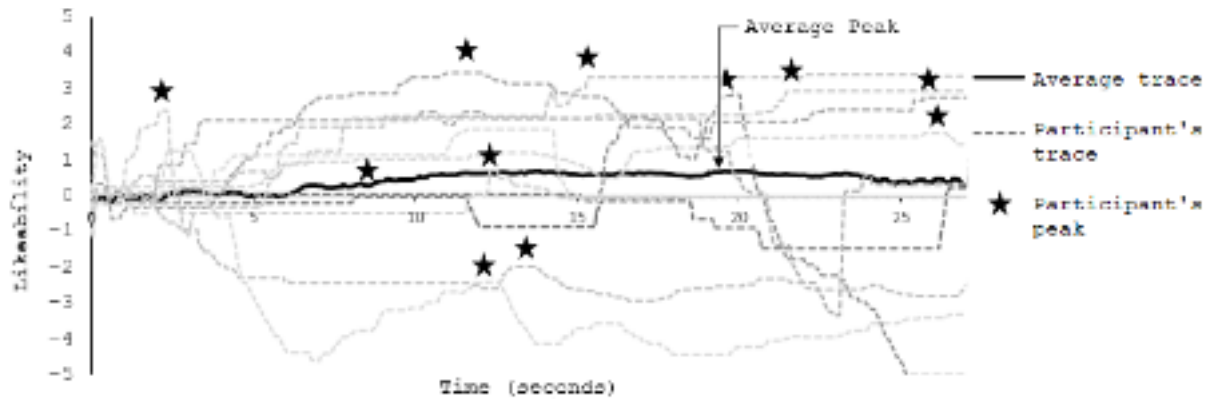
Figure 9. The differences in the location of the peak for the average trace and the participants' trace. For simplicity reasons, the trace of ten participants is shown, and out of those ten participants, only one person' location of the peak represented the peak of the average trace. Horizontal axis length of the commercial in seconds (0 to 40); vertical axis momentary likeability of the commercial (range -2 to 2). The continuous line shows the average affective trace of the participants. The dashed line represents the individual affective trace of a person. Star represents the peak experience of a person.

4.1.3 Regression model

Is it possible to make a regression model out of the combination of the ValueFlow proxy predictors? The results show that the factors correlate so high with each other that adding predictors to the model will not give any additional information. Therefore, it can be concluded that using a multiple regression model with ValueRate, peak- and final moment as predictors will not make a better prediction for OAL than the factors alone. This can be explained by the fact that the data of the ValueFlow does not actually represents the peaks and final moment of the individual participants, causing these key elements to lose their valuable information. The result being that no information is added to the regression model, above that of the ValueRate.

4.1.4 Familiarity check

The results show no interaction effect for familiarity. The findings reported above therefore hold when commercial familiarity is taken into account. Meaning that the likability of a commercial is

not influenced by whether a commercial is well-known. The results are thus not affected by not controlling for commercial familiarity.

4.2 Part 2: Transfer to other marketing measurements

The results of the exploratory study for Validators show strong relationships to the other marketing metrics, except for brand recall. As in the research of Baumgartner, similar results were found for the transfer to brand liking. A strong relationship between those variables was expected. One of the most important aspects of creating a good commercial is that for a person after seeing the ad a more positive feeling about the brand is created, which, as shown by the results, can be achieved by making a highly likeable commercial. The results found, contradictory to the results of Baumgartner et al. (1997) a weak relationship with brand recall. These differences can be explained by the fact that in Baumgartners' research all the commercials were positively assessed. The results of the current study show, as in the research of Baumgartner, that higher liked commercials are indeed better remembered, however, negatively rated commercials also show high scores on recall, causing the low correlation between those two variables. Medium-rated commercials are remembered least, which followed the idea that brand recall is not triggered by how much a commercial is liked, but how striking a commercial is (Smit et al., 2006). For both call-to-actions, a strong relationship was found. The findings that there is a strong link between those variables shows indeed that a well-liked commercial creates a sense of urgency that motivated to act, both in buying the product or trying to get more information. Last, the data of the ValueFlow also showed a strong relationship with commercial 'branding'. A well-liked commercial leads to a commercial with a strong link to the brand, but also the other way around, a commercial with a stronger link with a brand leads to a higher liked commercial.

Both the ValueRate and the final moment had the strongest relationship with the different marketing metrics, but the peak experience had a weaker relationship. While both the final moment and the peak experience are not actually representing the peak experience and final moment of the individual participants, the final moment is less prone to bias than the peak. This could be explained

by that the peak experience deviates both on the dimension of liking and on the dimension of time from the actual experience of a individual participant, while the final moment is only deviating on the dimension of liking. Therefore, the final moment is probably better representing the actual liking of a commercial and this could explain the higher correlations found for the final moment than the peak experience.

In short, for the ValueFlow a comparable transfer to the other marketing metrics as overall ad liking was found, reminding that OAL is one of the strongest factors predicting marketing success (Haley and Baldinger, 1991; Smit, Meurs & Neijens, 2006). The ValueFlow does measure the same concept but on a momentary level. Finding a similar result for the transfer of the ValueFlow towards other marketing measurement indicates further validation of the use of a ValueFlow as a valid marketing measurement tool.

4.2.1 Product interest check

The results show no interaction effect for product interest on call-to-action. Meaning that the intention to take action is not influenced by the interest a person has in the product regardless of the commercial. These results were not in line with previous research by Pelsmacker & Janssens (2007). They found that buying behaviour and information search behaviour are strongly affected by product interest. Therefor, not finding results led to doubts to what extent a person could make an estimation of their interest in product regardless of the commercial. More research is needed to determine whether a person is actually able to make an unbiased estimation of their interest when the commercial about that product has just been shown. Therefor, conclusions about whether product interest is not affecting the variance of the call-to-actions should be taken with caution.

4.3 Confound factors: ad duration & beginning of the measurement

Results showed that the SMAE is affected by the duration of a commercial and the ValueRate is not. These results confirm that Baumgartner et al. (1997) should not have used the sum of the momentary affective experience as a baseline for comparison with the variables peak experience and

final experience. Taking the sum of all experiences would mean that a longer commercial, would lead to more momentary points of experience and therefore lead to a proportionately higher SMAE than a shorter commercial. Applying a correction for the length of a commercial showed improvement in the model of predicting OAL out of the data of the ValueFlow and thereby justified the use of the variable ValueRate instead of SMAE as a baseline variable.

The analyses for the noisy beginning data showed that for ValueRate the beginning does indeed bring error to the model for predicting OAL. Deleting the average over the data from the first measurement until the first turning point resulted in significant improvement to the variable ValueRate and a significantly better prediction of OAL. These results lead to the conclusion that the data of the first 3.74 seconds are indeed less informative and should be removed.

4.4 Methodological issues

There are concerns that moment-to-moment measurements cause participants to unnatural introspection while they are observing the commercial. Therefore, creating an artificial self-awareness which could cause an abnormal viewing experience, and therefore fundamentally changing an advertising experience into a research experience (Micu, & Plummer, 2010). It is unknown to which extent this effect has played a role in this research.

Brewer and Lichtenstein (1981) explained that the presence of a storyline, which is present in most commercials, could cancel out the preference for patterns of affective response. The same accounts for the peak-end rule. Also, because commercials try to persuade people, a person has to respond to marketing influence attempts and therefore reactions to advertisements might differ from responses to other affective experiences (Friestad & Wright, 1994).

5. General conclusion

All in all, the results of part one of this study show, that in deriving an overall evaluation from the data of the ValueFlow, certain parts or specific key elements do not weigh more. Taking the average value over the affective responses across a commercial does evenly well predict the likeability of a commercial. For the ValueFlow, where for the analysis an average over the participants was taken, the peak-end rule did not hold. However, the peak-end rule, as described by Kahneman et al. (1993), has been studied for the integration of experiences of a single person. Further research is needed to determine whether for the ValueFlow the peak-end rule would have been robust if the analysis takes place on a personal level and to determine which extent it is valid to take the advertisement as the unit of analysis and treat the participants as replicates. Results of part two of this study indicate strong relationships between the data of the ValueFlow and the different marketing measurements. This is important, since all marketing measurements try in their own way to predict the same concept, namely marketing success. For the ValueFlow a comparable transfer was found as overall ad liking, indicating further validation of the use of ValueFlow as a valid marketing measurement tool.

References

Aaker, D. A., Stayman, D. M., & Hagerty, M. R. (1986). Warmth in advertising: Measurement, impact, and sequence effects. *Journal of consumer research*, *12*(4), 365-381.

Aaker, D. A., & Stayman, D. M. (1990). Measuring audience perceptions of commercials and relating them to ad impact. *Journal of Advertising Research*.

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, 34(2), 219-232.

Biel, A. L. (1990). Love the ad. Buy the product? *Admap*, *26*, 21-25.

Biel, A. L., & Bridgewater, C. A. (1990). Attributes of likeable television commercials. *Journal of advertising research*, *30*(3), 38-44.

Carmon, Z., & Kahneman, D. (1996). The experienced utility of queuing: real time affect and retrospective evaluations of simulated queues. Duke University: Durham, NC, USA.

Cook, C., Heath, F., Thompson, R. L., & Thompson, B. (2001). Score reliability in web or internet-based surveys: Unnumbered graphic rating scales versus Likert-type scales. *Educational and Psychological Measurement*, *61*(4), 697-706.

De Pelsmacker, P., & Janssens, W. (2007). A model for fair trade buying behaviour: The role of perceived quantity and quality of information and of product-specific attitudes. *Journal of business ethics*, *75*(4), 361-380.

Du Plessis, E. (1994a). Likeable ads work best, but what is 'likeability'. *Admap.(May)*, 10-13.

Du Plessis, E. (1994b). Understanding and using likability. *Journal of advertising research*, *34*(5), RC3- RC3.

Greene, W. F. (1992). Observations: what drives commercial liking? *Journal of Advertising Research*, *32*(2), 65-68.

Haley, R. I. (1990). The ARF Copy Research Validity Project: How it will add value to future
advertising decisions. *Transcript Proceedings of the Seventh Annual ARF Copy Research
Workshop. New York: Advertising Research Foundation*.

Haley, R. I., & Baldinger, A. L. (1991). The ARF copy research validity project. *Journal of
advertising research*.

Haley, R. I. (1994). A rejoinder to 'Conclusions from the ARF's Copy Research Validity
Project.'(response to article by John R. Rossiter and Geoff Eagleson, in this issue, p. 19).
Journal of Advertising Research, *34*(3), 33-35.

Haley, R. I., & Baldinger, A. L. (2000). The ARF copy research validity project. *Journal of
advertising research*, *40*(6), 114-135.

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is
preferred to less: Adding a better end. *Psychological science*, *4*(6), 401-405.

Klein, L. R. (1998). Evaluating the potential of interactive media through a new lens: Search versus
experience goods. *Journal of business research*, *41*(3), 195-203.

Leather, P., McKechnie, S., & Amirkhanian, M. (1994). The importance of likeability as a measure of
television advertising effectiveness. *International Journal of Advertising*, *13*(3), 265-280.

Leeflang, P. S., & Wittink, D. R. (2000). Building models for marketing decisions: Past, present and
future. *International journal of research in marketing*, *17*(2-3), 105-126.

Loewenstein, G. F., & Prelec, D. (1993). Preferences for sequences of outcomes. Psychological
review, 100(1), 91.

Loewenstein, G., & Sicherman, N. (1991). Do workers prefer increasing wage profiles?. Journal of
Labor Economics, 9(1), 67-84.

Micu, A. C., & Plummer, J. T. (2010). Measurable emotions: How television ads really work: Patterns
of reactions to commercials can demonstrate advertising effectiveness. *Journal of
Advertising Research*, *50*(2), 137-153.

Murdock Jr, Bennet B. "The serial position effect of free recall." *Journal of experimental psychology* 64.5 (1962): 482.

Polsfuss, M., & Hess, M. (1991). &Quot; Liking&Quot; Through Moment-To-Moment Evaluation; Identifying Key Selling Segments in Advertising. ACR North American Advances.

Redelmeier, D. A., & Kahneman, D. (1996). Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. Pain, 66(1), 3-8.

Ross, W. T., & Simonson, I. (1991). Evaluations of pairs of experiences: A preference for happy endings. *Journal of Behavioral Decision Making*, *4*(4), 273-282.

Rossiter, J. R., & Eagleson, G. (1994). Conclusions from the ARF's copy research validity project. *Journal of Advertising Research*, *34*(3), 19-33.

Shim, S., Eastlick, M. A., Lotz, S. L., & Warrington, P. (2001). An online prepurchase intentions model: The role of intention to search: Best Overall Paper Award—The Sixth Triennial AMS/ACRA Retailing Conference, 2000. *Journal of retailing*, *77*(3), 397-416.

Smit, E. G., Van Meurs, L., & Neijens, P. C. (2006). Effects of advertising likeability: A 10-year perspective. *Journal of Advertising Research*, *46*(1), 73-83.

Thorson, E., & Friestad, M. (1989). The effects of emotion on episodic memory for television commercials.

Walker, D., & Dubitsky, T. M. (1994). Why liking matters. *Journal of advertising research*, *34*(3), 9-19.

Young, C. E. (2008). The advertising research handbook. Ideas in Flight. 22-23.