

## Response latencies of pleasure and displeasure ratings: Further evidence for mixed feelings

Ulrich Schimmack

*University of Toronto, Mississauga, Canada*

All studies that examined the co-occurrence of pleasure and displeasure revealed at least some reports of mixed feelings (i.e., reports of concurrent pleasure and displeasure). Some researchers attribute these reports to measurement error, whereas others regard them as valid. This article examined response latencies of affect ratings to test the validity of reported mixed feelings. First, I demonstrate that respondents need more time to indicate the presence of an affect than the absence of an affect on unipolar response formats. Then, I demonstrate that mixed feelings are reported even when response latencies show a unipolar pattern for displeasure ratings, while the results for pleasure ratings were more mixed, but additional evidence suggests that they also are valid reflections of pleasure. In addition, I demonstrate that pleasure and displeasure ratings are independent of item-order and item-spacing. These results provide further support for the validity of reported mixed feelings and two-dimensional representations of pleasure and displeasure.

The relation between pleasure and displeasure is one of the oldest and most controversial questions in emotion research (Beebe-Center, 1932; Cacioppo & Berntson, 1994; Diener & Iran-Nejad, 1986; Larsen, McGraw, & Cacioppo, 2001; Larsen, McGraw, Mellers, & Cacioppo, 2004; Reisenzein, 1992; Rozin, 1999; Russell & Carroll, 1999; Schimmack, 2001). Some researchers consider pleasure and displeasure mutually exclusive, opposite ends of a single bipolar dimension (Russell & Carroll, 1999). Others consider pleasure and displeasure as distinct affects that vary along two separate dimensions (Cacioppo & Berntson, 1994; Diener & Iran-Nejad, 1986; Larsen et al., 2001; Schimmack, 2001).

Although many studies have tried to address this question empirically, these studies were unable to settle the debate for two reasons. First, Russell and Carroll

---

Correspondence concerning this article should be addressed to Ulrich Schimmack, Department of Psychology, University of Toronto, Mississauga, 3359 Mississauga Road North, Mississauga, Ontario, L5L 1C6, Canada; email: [uli.schimmack@utoronto.ca](mailto:uli.schimmack@utoronto.ca)

This research was funded by Social Sciences and Humanities Research Council Grant 410-2001-1560. I thank Agata Biazik, Priscilla Burnham, Nataine Clarke, Cynthia de Medeiros, Lucy Farhat, Susan Nguyen, Christine Rooks, Hana Tariq, and Christopher Tavares for their help.

(1999) pointed out two fundamentally different meanings of bipolarity (see also Schimmack, 2001). Bipolarity can be defined as a reciprocal relationship between pleasure and displeasure. That is, increases in pleasure imply decreases in displeasure and vice versa. Strong bipolarity would imply a perfectly negative correlation between intensity of pleasure and intensity of displeasure. For example, receiving a gift increases pleasure to the same extent as it decreases displeasure. Logical considerations and psychometric evidence contradicts this notion of bipolarity (Russell & Carroll, 1999; Schimmack, Bockenholt, & Reisenzein, 2002). Logically, increases in pleasure can occur without any accompanying changes in displeasure. For example, somebody who is in a good mood and does not feel displeasure can feel even better after winning \$1000 in a lottery without any changes in displeasure. Empirically, the correlation between pleasure and displeasure typically ranges from  $r = -.30$  to  $-.70$  (cf. Schimmack, 2001; Schimmack in press; Schimmack et al., 2002). Hence, pleasure and displeasure are not strictly independent, in the sense of varying independently of each other, but they are also not bipolar opposites, in the sense of an inverse relation between the intensities of pleasure and displeasure.

The present article examines the second notion of bipolarity. It defines bipolarity as mutual exclusiveness of two opposing entities. Accordingly, pleasure and displeasure are mutually exclusive feelings. When somebody feels pleasure, the individual cannot feel displeasure at the same time and vice versa. This notion of bipolarity has been tested by examining whether people can experience mixed feelings; that is, whether they have concurrent experiences of pleasure and displeasure (cf. Beebe-Center, 1932). Whereas numerous studies have examined the first notion of bipolarity, very few studies have examined the second notion of bipolarity that pleasure and displeasure are mutually exclusive feelings (Diener & Iran-Nejad, 1986; Larsen et al., 2001; Russell & Carroll, 1999; Schimmack, 2001).

Although all studies found some reports of mixed feelings, researchers disagree about the validity of these reports. In their seminal study, Diener and Iran-Nejad (1986) found frequent reports of pleasure and displeasure at low to moderate intensities, but few reports of intense pleasure and intense displeasure. The authors concluded that pleasure and displeasure are separate affects, which can co-occur at low to moderate intensities, but which are mutually exclusive at high intensities. Russell and Carroll (1999) also obtained concurrent reports of pleasure and displeasure. However, they concluded that reports of mixed feelings are invalid and are due to inappropriate items, random error, response styles, and misinterpretations of the response format. Schimmack (2001) proposed that pleasure and displeasure are reciprocally related, but they do not fully inhibit each other. As a result, pleasure and displeasure can co-occur when an increase in displeasure leads to a partial reduction in pleasure that does not fully eliminate pleasure. Larsen et al. (2001) also proposed that pleasure and displeasure are separate affects that can be experienced concurrently in situations

that elicit ambivalent emotions. However, most of the time experiences of pleasure and displeasure are mutually exclusive.

Evidently, the substantial question of whether pleasure and displeasure are mutually exclusive is intricately linked to the methodological question of the validity of self-reported feelings of pleasure and displeasure. This article makes a comprehensive attempt to examine the validity of reports of mixed feelings. For this purpose, it relies on the same procedure as the one used by Schimmack (2001). That is, systematic changes in affect due to a mood induction are used to rule out random error as an explanation, and ratings of hot and cold are used to detect individuals who misinterpret the response format. In addition, the present article uses two new methods to examine the validity of pleasure and displeasure ratings. First, response latencies are used to examine participants' interpretation of the response format. A pilot study demonstrates that response latencies yield different patterns for responses on a unipolar and a bipolar scale. Second, the order and the number of items between pleasure and displeasure ratings was manipulated to examine whether the time frame influences reports of mixed feelings.

### Misunderstanding of response formats

Meddis (1972) demonstrated that response formats influence respondents' ratings of pleasure and displeasure. In recent years, the reason for this effect of response formats has become evident (Gannon & Ostrom, 1996; Russell & Carroll, 1999; Schimmack et al., 2002). Participants tend to interpret symmetrical formats (e.g., Likert scales) as bipolar scales and asymmetric formats as unipolar scales. Take ratings of pleasure on a symmetric format (e.g., a 5-point Likert scale) and an asymmetric format (e.g., a 5-point intensity scale) as an example. On the Likert scale the lowest category implies a strong feeling of displeasure, the middle category implies the absence of pleasure and displeasure, and the highest category implies the presence of strong pleasure. The same response categories have different meanings on the intensity scale. The lowest category implies the absence of pleasure, the middle category implies moderate pleasure, and the highest category implies strong pleasure.

To examine whether pleasure and displeasure are mutually exclusive, it is essential to obtain separate reports of pleasure and displeasure on asymmetric, unipolar scales. However, asymmetric formats can produce misleading results when they are interpreted as bipolar scales (Russell & Carroll, 1999; Schimmack et al., 2002). In this case, participants may use the middle category to indicate that they felt neither pleasure nor displeasure. However, the researcher interprets this response as evidence for concurrent experiences of moderate pleasure and moderate displeasure. As a result, ratings on an asymmetric scale that is misinterpreted as a bipolar scale can provide false evidence for mixed feelings (cf. Russell & Carroll, 1999).

To address this concern, researchers need to communicate clearly to their participants the meaning of response categories. For example, some researchers have asked participants to make separate judgements of the presence and the intensity of affects (Ekman et al., 1987; Larsen et al., 2001; Reisenzein, 1995; Russell & Carroll, 1999). Intensity is rated only after participants report that an affect was present in the first rating. As a result, it is clear that low response categories on the intensity scale refer to low intensity of the present affect and not to strong intensities of an opposing affect.

One disadvantage of this method is that it doubles the number of items. To overcome this limitation, Schimmack (1997; Schimmack & Diener, 1997; Schimmack & Hartmann, 1997) proposed an alternative method. Respondents were instructed to respond to a single asymmetric response format in a two-step manner, that is, they were asked to first consider whether an affect is present or absent. Then they should respond with zero (i.e., the lowest response category) when an affect was absent, and they should use the remaining categories to indicate the intensity of an affect that was present. In addition, the zero category was visually removed from the remaining categories to emphasise the categorical difference between zero responses that indicate the absence of an affect, and other responses that indicate the presence of an affect.

Three studies obtained reports of mixed feelings with methods that ensured a unipolar interpretation of response formats (Larsen et al., 2001; Russell & Carroll, 1999; Schimmack, 2001). However, it remains possible that these reports were due to careless responding and that a few participants did not read or understand the instructions. Hence, it is important to provide additional evidence that reports of mixed feelings were not due to a misunderstanding of the response format.

The studies by Schimmack (2001) and Larsen et al. (2001) addressed this concern in several ways. Schimmack compared concurrent reports of pleasure and displeasure with concurrent reports of hot and cold on the same response format. If participants misinterpreted the response format as a bipolar scale, they should be equally likely to report concurrent feelings of hot and cold as feelings of pleasure and displeasure. However, participants were more likely to report concurrent feelings of pleasure and displeasure than feeling hot and cold. This method of validating reports of mixed feelings assumes, however, that participants used the response format consistently across items. Schimmack (2001) also observed that more participants reported mixed feelings after an induction of displeasure. This finding suggests that reports of mixed feelings after the mood induction were valid. This leaves, however, the possibility that reports of mixed feelings before the mood induction were invalid, and that reports of mixed feelings afterwards may have been due to demand effects.

Larsen et al. (2001) also relied on a comparison of affect reports in different situations. For example, significantly more participants reported mixed feelings of happiness and sadness after seeing the movie *Life is Beautiful* than before

seeing the movie. Again, this approach validates the reports of mixed feelings in special situations, but it cannot determine the validity of the reported mixed feelings in everyday situations. Subsequently, I propose that response latencies of affect ratings can be used to test whether reports of mixed feelings are valid.

### Response latencies of affect ratings

Schimmack (1997) proposed that response latencies of affect ratings reveal participants' interpretations of response formats (see also Schimmack & Hartmann, 1997). On a unipolar scale it should be easier to report the absence of an affect than the presence of an affect. The reason is that the absence of an affect is associated with a single response option (i.e., zero). In contrast, the presence of an affect is associated with multiple response options that reflect different intensities of an affect. Choosing the appropriate response category should take additional time. This prediction was supported for mood ratings (Schimmack, 1997) and for emotion judgements in response to hypothetical scenarios (Schimmack & Hartmann, 1997). The present study used the response latencies of affect ratings to examine the validity of concurrent reports of mixed feelings. Valid reports of mixed feelings imply that participants need more time to report the presence than the absence of displeasure. Participants should also take more time to report the presence than the absence of pleasure when displeasure is present. However, this prediction is more difficult to test because few participants report the absence of pleasure (Schimmack, 2001).

### Item-order and Item-spacing

Another unexplored issue is the influence of item-order on pleasure and displeasure ratings. Ito, Lang, and Cacioppo (1998) were concerned that pleasure and displeasure ratings that follow each other immediately create a demand effect to report only one of the two affects. To avoid this potential problem, they obtained pleasure and displeasure ratings of affective pictures in separate trials. However, no study has systematically examined whether pleasure ratings are indeed influenced by preceding displeasure ratings and vice versa. It is also unknown whether item-order effects are limited to ratings that follow each other or if they occur over longer intervals of a few items.

The influence of item-spacing on pleasure and displeasure ratings is also important for another reason, that is, it is self-evident that people can experience pleasure and displeasure over extended periods of time (Beebe-Center, 1932). Hence, the hypothesis of mutually exclusive feelings only makes sense for short time intervals (Russell & Carroll, 1999). As affect ratings are typically made sequentially, it is possible that some reports of mixed feelings are due to true changes in affect during an extended time interval. If this were the case, reports of mixed feelings should increase with the number of items between pleasure and displeasure ratings. This study examines this possibility.

## OVERVIEW

The first study validates the ability of response latencies to differentiate between responses on unipolar and bipolar response formats. The data were obtained in pilot ratings of affective pictures. One study asked for responses on a bipolar 7-point scale, ranging from  $-3$  = strong displeasure to  $+3$  = strong pleasure. The other study asked for two unipolar judgements of pleasure and displeasure on a 7-point scale ranging from  $0$  = "not at all" to  $6$  = "maximum intensity". Responses on the lowest category ( $-3$  and  $0$ , respectively) should be faster than other responses on the unipolar format, but not on the bipolar format because the lowest category on the unipolar format represents the absence of an affect.

The second study closely follows the study by Schimmack (2001). Participants first reported their momentary feelings before an experiment by means of ratings of 14 adjectives on a 7-point unipolar scale. Then they saw a series of appealing and appalling pictures as part of different experiments (these data are reported elsewhere). After the experiments, participants' momentary feelings were assessed with the same procedure as before the experiment. Schimmack (2001) used a paper-pencil questionnaire for the mood assessment. In contrast, the present study used computers to assess mood before and after the experiment. As a result, it was possible to present the items in a new random order for all participants to examine item-order and item-spacing effects. In addition, the computer could record the response latencies of the affect ratings.

## PILOT STUDY

The sole purpose of the pilot study was to demonstrate that response latencies of responses on a 7-point scale depend on the interpretation of the response format. Unipolar interpretations lead to faster responses on the lowest response category than on the other response categories (Schimmack, 1997). Response latencies of affect ratings on bipolar formats have not been examined, but they should not produce much faster responses on the lowest response category. Rather, faster responses are to be expected for the middle category, which represents the absence of pleasure and displeasure.

### Procedure

*Participants.* A total of 74 students at the University of Toronto, Mississauga participated in the study for course credit: 38 students made ratings on a bipolar format and 36 students made ratings on the unipolar format.

*Materials and procedure.* The study was run in groups of 1–5 students. Participants first signed an informed consent form, before they were seated in front of a computer. On the computer, participants first reported their current feelings before they proceeded to the main experiment. In the experiment,

participants saw emotional pictures (e.g., sunsets, cute animals, trash, dead animals) for 2 seconds. After each picture presentation, a 7-point rating scale appeared on the screen. Participants reported their affective reaction to the picture by moving the mouse to the appropriate response option and pressing the left mouse button. Participants' responses automatically triggered the presentation of the next picture. The pictures were presented in a new random order for each participant.

Participants in the bipolar condition made a single judgement on a 7-point scale. The response options were labelled,  $-3$  = strong displeasure,  $-2$  = moderate displeasure,  $-1$  = mild displeasure,  $0$  = neutral,  $+1$  = mild pleasure,  $+2$  = moderate pleasure,  $+3$  = strong pleasure. Participants in the unipolar condition made one judgement for pleasure and another judgement for displeasure. The order of the two judgements varied randomly from picture to picture. Judgements were made on a 7-point scale, with the response options:  $0$  = no, not at all,  $1$  = yes, very mild,  $2$  = yes, mild,  $3$  = yes, moderate,  $4$  = yes, strong,  $5$  = yes, very strong, and  $6$  = yes, maximum intensity. The no/yes part of the labels reinforces the distinction between the lowest response option, indicating the absence of an affect, and the other options, indicating the presence of an affect. In addition, participants received the standard instructions about the unipolar nature of the response format (Schimmack, 1997; Schimmack & Diener, 1997).

## Results

It is common for response latencies of judgements to be skewed and to include extreme outliers because participants take a break or are distracted. To remove these artifacts from the data, trials with extremely short (less than 500 ms) and long (more than 10 s) response latencies were removed. Response latencies that were more than two standard deviations above or below each individual's mean response latency were also removed, leaving 93% of the response latencies to be analysed. Finally, a logarithmic transformation was used to normalise the response latencies.

An analysis of variance (ANOVA) is inappropriate for the analysis of these data because participants chose response options with different frequencies. Rather, I used hierarchical linear and nonlinear modelling (HLM 5) to test statistical significance (Raudenbush, Bryk, Cheong, & Congdon, 2000). HLM 5 provides estimates of the average latencies for the seven response categories on the two response formats (Figure 1). Figure 1 shows that participants needed more time to respond to the unipolar format than the bipolar format. A significant effect for polarity on the Intercept in the HLM model shows that this difference is statistically reliable (Table 1). More importantly, Figure 1 confirms that response formats produce different patterns of response latencies. On the unipolar format, responses to the lowest category are faster than responses to the

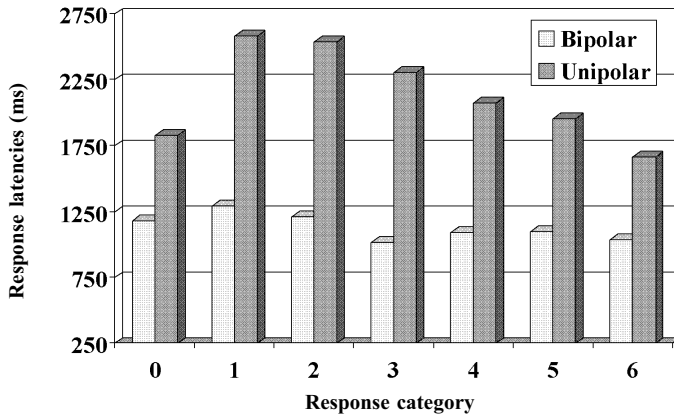


Figure 1. Response latencies in the pilot study.

TABLE 1  
Coefficients of the hierarchical linear and nonlinear model (HLM) in the pilot study

<i>Effect</i>	<i>Coefficient</i>	<i>Standard error</i>	<i>t-ratio</i>
Intercept	7.08	.03	203.53
Polarity	0.43	.05	8.48
<i>Dummy 1</i>			
Intercept	0.09	.03	3.22
Polarity	0.25	.04	7.37
<i>Dummy 2</i>			
Intercept	0.01	.03	0.43 (n.s.)
Polarity	0.31	.04	7.37
<i>Dummy 3</i>			
Intercept	-0.16	.03	6.01
Polarity	0.39	.04	8.92
<i>Dummy 4</i>			
Intercept	-0.09	.03	3.10
Polarity	0.21	.05	4.52
<i>Dummy 5</i>			
Intercept	-0.08	.03	3.17
Polarity	0.15	.05	3.35
<i>Dummy 6</i>			
Intercept	-0.13	.03	3.86
Polarity	0.03	.05	0.59 (n.s.)

Polarity: 0 = Bipolar, 1 = Unipolar; *Dummy 1*: Bipolar (-3 vs. -2) Unipolar (0 vs. 1), *Dummy 2*: Bipolar (-3 vs. -1) Unipolar (0 vs. 2); *Dummy 3*: Bipolar (-3 vs. 0) Unipolar (0 vs. 3); *Dummy 4*: Bipolar (-3 vs. +1) Unipolar (0 vs. 4); *Dummy 5*: Bipolar (-3 vs. +2) Unipolar (0 vs. 5); *Dummy 6*: Bipolar (-3 vs. +3) Unipolar (0 vs. 6).

other categories with the exception of the highest category. In contrast, the bipolar format produced relatively homogeneous response latencies for the seven response categories. The HLM model tests the statistical significance of the differences between the two patterns with the interaction terms between the Level 2 variable *format* and the Level 1 dummy variables representing the response categories. All interaction terms except the one for the highest response category were significant.

An inspection of Figure 1 suggests that responses on the bipolar scale were the most rapid in the middle category. Additional significance tests with an HLM model limited to the bipolar reaction times and the middle category as the comparison category revealed that responses to the middle category were significantly faster than responses to all but the highest response category,  $t(37) > 3$ . This finding is consistent with the idea that participants are generally faster to report the absence of a feeling even on a bipolar scale. The unipolar format shows that response latencies decrease with increasing levels of intensity. This pattern has been found in earlier studies (Schimmack, 1997; Schimmack & Diener, 1997; Schimmack & Hartmann, 1997). Seemingly, participants are faster in reporting intense experiences of affect but need more time to detect or discriminate weak affects.

## Discussion

The pilot study examined the ability of response latencies to discriminate bipolar interpretations from unipolar interpretations of a response format. On a unipolar format the lowest response category has a special status because it is the only category that reflects the absence of an affect. Hence, participants can choose this option as soon as they have determined that an affect is absent. In contrast, all the other categories reflect the presence of an affect and require participants to choose the option that best matches the intensity of their experience. As a result, these responses take more time. On bipolar formats the lowest category is not categorically different from other categories, and response latencies for the lowest category should not differ categorically from response latencies to other response options. This prediction was supported by the data. One minor exception appears to be the highest response category, which did not differ between conditions (see Gannon & Ostrom, 1996, for similar results). The most plausible explanation for this finding is the clarity of extremely intense experiences. If an experience is extremely strong, participants do not have to choose among different categories indicating the presence of an affect. Rather, they can quickly choose the highest response category. The fast responses to the most extreme response category do not diminish the practical utility of response latencies because the highest response category is chosen quite rarely, especially during the assessment of affect in nonemotional situations. In sum, faster responses to the lowest response category than to other response categories

reflect unipolar interpretations of a response format. This finding provides the empirical foundation for the main study.

## MAIN STUDY

The main study used response latencies of affect ratings before and after a mood induction to examine the validity of reports of mixed feelings. Schimmack (2001) found relatively few and weak reports of mixed feelings in a neutral situation before an experimental mood induction. The reason for this finding was that most participants reported feeling pleasant without feeling unpleasant. A few remaining participants reported pleasure and displeasure. However, it is possible that these participants did not really experience mixed feelings, but rather used the response format as a bipolar scale. After the negative mood induction, the frequency and intensity of mixed feelings increased because more participants reported displeasure while still reporting pleasure. The increase in mixed feelings suggests that participants felt mixed feelings after the experiment. However, this increase may have been due to demand effects. In the present study, I used the response latencies of pleasure and displeasure ratings to examine the validity of reports of mixed feelings before and after the experiment. If participants correctly use the intensity scale as a unipolar format, then participants should need more time to report the presence of displeasure than the absence of displeasure. This should even be true when participants are reporting pleasure.

The main study also manipulated the item-order and the item-spacing (i.e., the number of items between pleasure and displeasure ratings). Hence, I was able to examine whether people would be more likely to report mixed feelings of pleasure and displeasure over longer periods of time; that is, when pleasure and displeasure ratings were separated by several other items.

As the main study is very similar to the original study by Schimmack (2001), I was also able to replicate key findings in the earlier publication, namely: (a) seeing positive and negative affective pictures increases displeasure and decreases pleasure; (b) participants report more mixed feelings of pleasure and displeasure than concurrent feelings of hot and cold; and (c) participants report more intense mixed feelings after a negative mood induction than in a neutral situation before a mood induction.

## Method

*Participants.* A total of 1118 students at the University of Toronto, Mississauga, participated in this study for course credit (816 female, 302 male). The data were collected as part of an ongoing study of affective reactions to affective pictures. Participants in all studies saw appealing and appalling pictures, although they made different types of ratings and saw different types of pictures. However, for the present study data were collapsed across experimental conditions.

*Materials and procedure.* The experiment was conducted in groups of one to five students. Participants were seated in front of a computer. They first answered a few demographic questions. Then they read the instructions for the first mood assessment: "If you do not experience the feeling, please respond with 0. If you do experience the feeling, please respond with 1 to 6, depending on the intensity of the feeling". Afterwards, they clicked a button to start the mood-rating task. The computer displayed 14 items in a new random order for each participant (pleasant, unpleasant, awake, tired, tense, calm, interested, bored, excited, conflicted, hot, cold, hungry, thirsty). The screen showed the sentence "Right now I feel..." followed by one of the 14 adjectives. Underneath were seven response buttons, which were labelled "0 = no, not at all", "1 = yes, very mild", "2 = yes, mild", "3 = yes, moderate", "4 = yes, strong", "5 = yes, very strong", "6 = yes, maximum intensity". Participants responded with a mouse click on one of the seven response buttons. After the mood assessment, participants were exposed to a series of appealing and appalling pictures that elicited pleasure and displeasure. Immediately after the last picture, the computer administered the second mood assessment.

## Results

*Preliminary analyses.* Two participants were eliminated because they had zero variability in their mood ratings before and after the main task, suggesting that they did not provide valid mood reports.

Ratings of feeling hot and cold were included as artifact measures (Schimmack, 2001). Under normal circumstances hot and cold are mutually exclusive feelings and should conform to the predictions of a unidimensional model. Hence, hot and cold ratings can be used to examine the influence of measurement artifact on reports of mixed feelings. I first examined hot and cold ratings before and after the experiment. Hot ratings decreased from Time 1 ( $M = 1.08$ ,  $SD = 1.37$ ) to Time 2 ( $M = 0.88$ ,  $SD = 1.30$ ),  $t(1115) = 5.39$ ,  $p < .01$ . Cold ratings also decreased from Time 1 ( $M = 0.67$ ,  $SD = 1.17$ ) to Time 2 ( $M = 0.50$ ,  $SD = 1.02$ ),  $t(1115) = 5.22$ ,  $p < .01$ . When cold-ratings were subtracted from hot ratings, temperature did not change from Time 1 ( $M = 0.41$ ,  $SD = 1.84$ ) to Time 2 ( $M = 0.38$ ,  $SD = 1.68$ ). This pattern of results suggests that most participants made unipolar hot and cold ratings. The bipolar difference-scores show that most participants felt neither cold nor hot. If participants had responded to the unipolar hot and cold items in a bipolar manner, they would have used the middle response category to report feeling neither hot nor cold. However, the means of hot and cold ratings were close to zero, which is consistent with a unipolar interpretation of the response format.

The next analyses examine reports of feeling hot and cold, which are assumed to reflect measurement error. I computed the percentage of participants who reported feeling hot (hot ratings greater than zero) and cold (cold ratings greater

than zero) (cf. Larsen et al., 2001; Schimmack, 2001). The vast majority of reports were consistent with a unidimensional model of hot and cold. At Time 1, 14% of the participants reported feeling hot and cold. At Time 2, 12% of the participants reported feeling hot and cold. The difference between the percentages was not significant, chi-square ( $df = 1116$ ) = 2.94,  $p = .09$ . A total of 901 participants (81%) responded to the hot and cold items at two separate occasions in a manner that is consistent with a unipolar interpretation of the response format; at least for the items hot and cold. The following analyses focus on this subsample of participants.

The response latencies of hot and cold ratings of the 901 participants who made unipolar and mutually exclusive hot and cold ratings should conform to the typical pattern of a unipolar scale; that is, faster response latencies for zero ratings than for non-zero ratings. As in the pilot study, response latencies were log-transformed to normalise distributions for statistical analysis. Figure 2 shows the retransformed mean latencies for the seven response categories. The pattern in Figure 2 conforms to the pattern for the unipolar response format in the pilot study. Statistical significance was assessed by means of regression analyses. For this purpose, I created dummy variables that represented the direct contrast between the zero category and the other categories. For example, for hot-ratings before the experiment I created six dummy variables that contrasted participants who responded with zero from participants who responded with 1, 2, 3, 4, 5, or 6. It has to be noted that these analyses have relatively little power for the highest response categories (5,6), which were used infrequently. All four regression analyses revealed a significant overall effect,  $F_s > 13$ ,  $p_s < .01$ . Hot and cold ratings before the experiment revealed significant contrasts for categories 1,2, and 3. Hot and cold ratings after the experiment revealed significant contrasts for categories 1,2,3, and 4. Overall, these results confirm that participants made hot and cold ratings in a unipolar

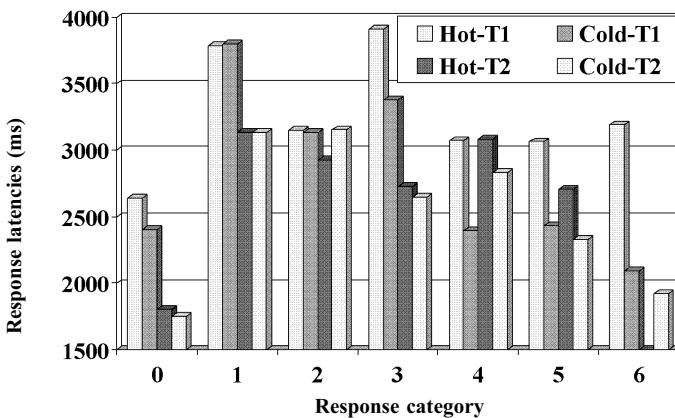


Figure 2. Response latencies of hot and cold ratings in the main study.

fashion. Furthermore, the analyses reveal that it is possible to find the characteristic unipolar pattern of response latencies in analyses of a single item across participants. Hence, it is possible to use the response latencies of pleasure and displeasure ratings to determine how participants responded to these items.

*Pleasure and displeasure.* Consistent with numerous studies, participants reported moderate pleasure ( $M = 3.04$ ,  $SD = 1.31$ ) and very low levels of displeasure ( $M = 0.48$ ,  $SD = 0.96$ ) at Time 1, in the neutral situation before the experiment. At Time 2, pleasure was significantly lower ( $M = 1.91$ ,  $SD = 1.33$ ),  $t(900) = 23.50$ ,  $p < .01$ , and displeasure had significantly increased ( $M = 1.53$ ,  $SD = 1.57$ ),  $t(900) = 18.41$ ,  $p < .01$ . This finding replicates Schimmack's (2001) finding that a mild negative mood induction increases displeasure and decreases pleasure.

Following Schimmack (2001), I used the MIN statistic to quantify the intensity of mixed feelings of pleasure and displeasure. The MIN statistic has a long tradition in attitude research to measure attitudinal ambivalence (Kaplan, 1972; Priester & Petty, 1996), and it provides a direct test of the co-occurrence of two affects (cf. Schimmack, 2001). For ratings of pleasure and displeasure, MIN assumes the value of the lower rating. For example, if a participant reports feeling neither pleasure ( $P = 0$ ) nor displeasure ( $D = 0$ ),  $MIN[P,D]$  indicates that the participant does not feel mixed feelings of pleasure and displeasure  $MIN[P,D] = 0$ . Even when a participant reports feeling extreme pleasure ( $P = 6$ ) without displeasure ( $D = 0$ ),  $MIN[P,D]$  indicates that this individual does not feel mixed feelings as MIN assumes the value of the lower displeasure rating  $MIN[P,D] = 0$ . However, if an individual reports feeling extreme pleasure ( $P = 6$ ) and mild displeasure ( $D = 2$ ),  $MIN[P,D]$  would indicate that this individual has mixed feelings of mild intensity ( $MIN = 2$ ).  $MIN[P,D]$  assumes the highest intensity of mixed feelings when pleasure and displeasure are at maximum levels,  $P = 6$ ,  $D = 6$ ,  $MIN[P,D] = 6$ . Hence, MIN captures all the aspects that are essential for the assessment of mixed feelings. It remains 0 as long as reports of two affects are mutually exclusive so that the intensity of one affect is 0, when the other affect is experienced. It increases in intensity as the intensity of concurrent feelings increase and it reaches its maximum when both affects are experienced with maximum intensity.

At both times, the intensity of mixed feelings (MIN) was significantly greater than zero,  $F_s > 221$ ,  $p_s < .01$ . As the participants in this analysis had  $MIN[hot,cold]$  scores of zero, this finding confirms earlier findings that participants are more likely to report feeling pleasant and unpleasant than hot and cold (Schimmack, 2001). Also consistent with earlier findings, the intensity of mixed feelings increased from Time 1 ( $M = 0.36$ ,  $SD = 0.68$ ) to Time 2 ( $M = 0.72$ ,  $SD = 0.87$ ),  $t(900) = 11.27$ ,  $p < .01$ . At Time 1, 24% of the participants reported mixed feelings and 51% of the participants reported mixed feelings at Time 2.

The following analyses examine the response latencies of pleasure and displeasure ratings. The results for displeasure ratings are presented first. Figure 3 shows the retransformed latencies for the seven response categories. The pattern is consistent with previous patterns for unipolar responses. Regression analyses showed overall effects of response category at Time 1 and Time 2,  $F_s > 16$ ,  $p_s < .01$ . At Time 1, contrasts for categories 1, 2, 3, and 4 were significant. At Time 2, contrasts for categories 1, 2, 3, 4, and 5 were significant. This finding shows that response latencies of displeasure ratings show the expected unipolar pattern.

The analyses of pleasure ratings before the experiment has to be interpreted with caution because only 42 participants (5%) reported the absence of pleasure ( $P = 0$ ). The pattern of pleasure ratings before the experiment does not show the characteristically faster responses to the lowest response category (Figure 4).

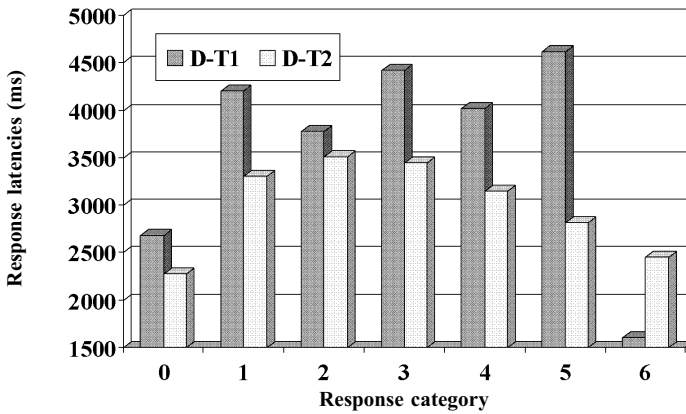


Figure 3. Response latencies of displeasure rating in the main study.

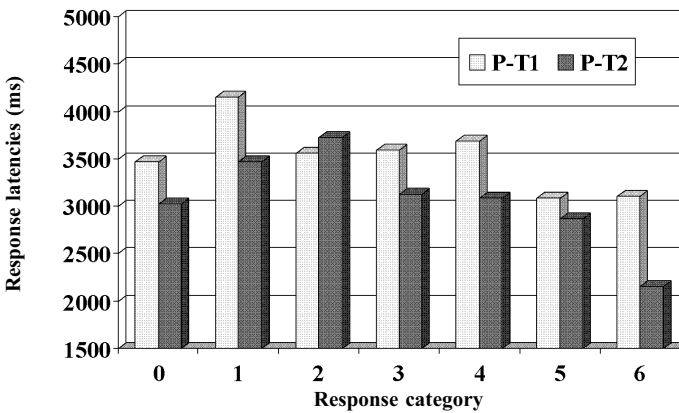


Figure 4. Response latencies of pleasure rating in the main study.

While the overall test of the regression analysis for Time 1 was significant,  $F(1, 900) = 2.32$ ,  $p = .03$ , none of the individual contrasts was significant. Consistent with the decrease in mean levels of pleasure, more participants reported the absence of pleasure at Time 2 ( $N = 156$ , 17%). At Time 2, the regression analysis revealed a significant overall effect,  $F(1, 900) = 5.07$ , and the contrasts for categories 1 and 2 were significant. This finding shows that the answer to the validity of mixed feelings rests on the pleasure ratings. Either pleasure is qualitatively different from other affects in that its absence is not a salient and easily reportable state, or participants use the asymmetric response format differently to rate pleasure. These alternative interpretations are explored in detail in the discussion.

*Item-order and item-spacing.* Affect ratings before and after the experiment were analysed separately. I determined the order of the two judgements (0 = pleasure first, 1 = displeasure first), and the number of items between the ratings of pleasure and displeasure (ranging from 0 to 12). The two predictors were correlated with pleasure, displeasure, and mixed feelings, MIN[P,D]. At Time 1, the magnitude of all correlations was smaller than .10, and none of the correlations was significant. The same finding was obtained at Time 2. In sum, the results show that pleasure and displeasure ratings are not influenced by the order or proximity of the two ratings. This finding implies that participants were just as likely to report mixed feelings when pleasure and displeasure ratings immediately followed each other as when the two items were separated by numerous items. Although it is difficult to prove that these factors have no influence at all, the small amount of explained variance and the large sample size suggest that item order and item spacing are negligible factors in the assessment of pleasure and displeasure.

## DISCUSSION

The main study replicated and extended previous findings by Schimmack (2001). Over 1000 participants reported their current feelings before and after various experimental tasks. Ratings of feeling hot and cold were used to detect individuals who clearly understood the unipolar nature of the response format. Over 80% of the participants made hot and cold ratings in a manner that was consistent with a unipolar response format. Furthermore, response latencies of these participants' hot and cold ratings conformed to the typical pattern of response latencies for unipolar formats. Pleasure and displeasure ratings of these participants revealed that they reported mixed feelings of pleasure and displeasure before and after the experimental tasks. Furthermore, the intensity of mixed feelings was stronger after the experimental tasks due to a significant increase in displeasure. Response latencies of pleasure and displeasure ratings were used to examine whether participants made pleasure and displeasure ratings in a unipolar manner. The data clearly confirmed that displeasure ratings

were made in a unipolar manner. Hence, these reports cannot be attributed to a misunderstanding of the response format for displeasure ratings. However, the results for response latencies of pleasure ratings were more ambiguous and it remains possible that participants responded to this item in a bipolar manner. Hence, it remains possible that reports of mixed feelings are invalid.

The present study also examined whether pleasure and displeasure ratings are influenced by item-order. Conflicting claims exist in the literature. Some researchers argue that pleasure and displeasure ratings that immediately follow each other create a demand effect to respond in a bipolar manner. Other researchers have argued that ratings of pleasure and displeasure that are not made in close temporal proximity can produce misleading evidence for mixed feelings because feelings can change rapidly from one moment to the next. As a result, participants may report pleasure at one time and displeasure a few moments later, even though they do not experience pleasure and displeasure at the same time. Contrary to these hypotheses, the present study failed to find significant item-order or item-spacing effects in a large sample of 901 participants. Hence, it is not possible to attribute the reciprocal relation between pleasure and displeasure to demand effects of pleasure and displeasure items that immediately follow each other. It is also not possible to attribute reports of mixed feelings to extended time periods that allow for fluctuation in current affect. This conclusion is also supported by a recent study of mixed feelings in response to disappointing wins (Larsen et al., 2004). Participants who could win \$5 or \$12 and actually won \$5 reported more mixed feelings than participants who actually won \$12. To examine the co-occurrence of pleasure and displeasure, participants were asked to press separate buttons for pleasure and displeasure. Participants pressed both buttons concurrently for extended periods of time and did not alternate between responses. This finding corroborates the present findings of concurrent rather than alternating experiences of pleasure and displeasure.

In sum, the present study provides an extensive and comprehensive validation study of pleasure and displeasure ratings. As noted in the introduction, the substantial question of the nature of pleasure and displeasure is intricately linked to the methodological question of validity. All studies that have examined the co-occurrence of pleasure and displeasure have found reports of mixed feelings. However, these studies were unable to settle the debate about the dimensionality of pleasure and displeasure because researchers disagreed about the validity of these reports. In particular, it was argued that evidence for two-dimensional models of pleasure and displeasure was due to: (a) selection of inappropriate items; (b) imprecise hypothesis; (c) inappropriate statistics; (d) random measurement error; (e) misinterpretations of response formats; and (f) unclear specification of time-frames.

The present study addressed all of these concerns. I used the item labels *pleasant* and *unpleasant* as clear markers of valence. I defined bipolarity clearly

as mutually exclusive experiences. I used the MIN statistic as an appropriate statistical test of mutual exclusivity. I demonstrated systematic changes in mixed feelings to rule out random measurement error as an explanation. I used hot and cold ratings and response latencies to examine the interpretation of the response format. I demonstrated that reports of mixed feelings are as likely to occur within the 4 seconds that it takes to answer two questions as within the 30 seconds that it took participants to respond to 14 items.

### Validity of pleasure ratings

The results of this study demonstrated that the validity of reports of mixed feelings, and therewith the validity of two-dimensional models of pleasure and displeasure, rests on the interpretation of pleasure ratings. A bipolar interpretation of the pleasure item is consistent with a one-dimensional model of pleasure and displeasure. A unipolar interpretation of the pleasure item is consistent with two-dimensional model of pleasure and displeasure. The two interpretations have radically different implications for the interpretation of affect ratings before and after the mood induction in the present study.

If participants responded to the pleasure item in a bipolar manner, then their ratings before the experiment did not reveal moderate feelings of pleasure. Rather, a rating of 3 indicates that they feel neither pleasant nor unpleasant. This is also consistent with a low rating on the unipolar displeasure item. A rating of 1 or 2 after the mood induction on the pleasure item indicates that the mood induction induced displeasure. This is also consistent with an increase in displeasure on the unipolar displeasure item. As a result, a bipolar interpretation implies that mood changed from neutral to negative.

On the other hand, a unipolar interpretation of pleasure implies that ratings of 3 on pleasure reveal moderate pleasure. In combination with the low displeasure ratings, this pattern suggests a positive mood. After the mood induction pleasure decreased and displeasure increased to about equal levels. Hence, the hedonic balance of pleasure and displeasure has moved from positive to neutral.

Fortunately, it is relatively easy to test these alternative accounts of the data. One only needs to assess pleasure and displeasure with a bipolar scale. If ratings of moderate pleasure merely reflect neutral experiences, then bipolar ratings before the experiment should be neutral. After all, this is what participants are doing when they check the middle category on a pleasure item, which is interpreted as a bipolar item. However, if ratings on an explicit bipolar item are above the middle category, then participants are trying to communicate that they are feeling good even in a rather neutral situation, implying that the moderate pleasure ratings on a unipolar scale are valid reports of moderate pleasure. Numerous studies have assessed pleasure-displeasure with the explicit bipolar Affect Grid (Russell, Weiss, & Mendelsohn, 1989). The scale ranges from 1 = extreme displeasure to 9 = extreme pleasure. On this scale ratings of 5 indicate

neutral states of feeling neither pleasant nor unpleasant. In the original study, ratings of current mood had a mean score of 5.96. Similar results were obtained in a recent study by Moskowitz, Pinard, Zuroff, Annable, and Young (2001), who used the Affect Grid to assess everyday moods. Schimmack et al. (2002) obtained pleasure and displeasure ratings on a 5-point bipolar Likert scale. Although the mean levels were not reported, I reexamined the data and obtained a mean rating of 3.46, which was significantly higher than the neutral response (3),  $F(1, 258) = 72.97, p < .01$ .

In sum, numerous studies with bipolar rating scales have found that participants report feeling pleasure rather than being in a neutral state in situations similar to the one in the present study before the experiments. Hence, the reports of moderate level of pleasure in the present study cannot be attributed to bipolar interpretations of the pleasure item. As a result, the overall pattern of results is most compatible with a model of pleasure and displeasure as separate, yet reciprocally activated feelings. In neutral situations, pleasure is at moderate levels and displeasure is often absent. Negative events induce displeasure and displeasure increases with the intensity of the negative event. At the same time, negative events reduce pleasure. As a result, negative events tend to induce mixed feelings because mild to moderate negative events do not fully suppress pleasure. However, due to the reciprocal relation between pleasure and displeasure it is uncommon for pleasure and displeasure to be experienced concurrently at high levels of intensity.

### Implications for the prevalence of mixed feelings

Proponents of two-dimensional models of pleasure and displeasure do not fully agree about the practical significance of mixed feelings. Larsen et al. (2001) caution that mixed feelings may occur only in rare situations. In contrast, Schimmack (2001) proposed that mixed feelings are quite common, and that most experiences of displeasure are accompanied by experiences of pleasure. The reason is that most experiences of displeasure are too mild to fully suppress pleasure. Fortunately, it is possible to reconcile these two positions. Larsen et al.'s (2001) position is correct in terms of the absolute frequency of mixed feelings. The authors found that 10% to 20% of their participants reported mixed feelings in typical everyday situations. In the present study, 24% of participants reported mixed feelings in a neutral situation and the intensity of mixed feelings was close to zero. Hence, one could argue that researchers can safely ignore mixed feelings in these situations.

However, the low intensity and frequency of mixed feelings in these situations has to be considered in the context of the low frequency and intensity of displeasure. In the present study, only 28% of the participants reported feeling displeasure before the mood induction, and the intensity of displeasure was only 0.48 on an intensity scale ranging from 0 to 6. This intensity is only slightly higher than the intensity of mixed feelings. Indeed, before the experiment the

intensity of displeasure and the intensity of mixed feelings are highly correlated ( $r = .79$ ) because most participants reported more intense pleasure than displeasure. As a result, displeasure is the major determinant of the intensity of mixed feelings, which is defined as the intensity of the weaker affect. It follows that ignoring mixed feelings essentially implies ignoring displeasure. Few researchers would be willing to ignore displeasure in affect assessment just because it is a relatively infrequent experience.

Furthermore, reports of mixed feelings are actually quite frequent compared to frequencies of other affective experiences. For example, Shapiro, Jamner, Goldstein, and Delfino (2001) found that participants in an experience sampling study reported anger only 6% of the time. Nobody would suggest that the low absolute frequency of anger implies that emotion theories can ignore anger. Similarly, we cannot ignore mixed feelings just because people do not experience mixed feelings all of the time.

## Outlook

This study alone cannot resolve the controversy surrounding the nature and dimensionality of pleasure and displeasure. However, it contributes to this controversy by providing further evidence for the validity of concurrent reports of pleasure and displeasure even in a typical assessment situation without an experimental manipulation of affect. At present there exists no evidence that all of these reports or even the majority of them are invalid. Hence, it is appropriate and meaningful to assess pleasure and displeasure separately. A premature assessment of pleasure and displeasure along a single bipolar dimension would preclude future examinations of the separate causes and consequences of pleasure and displeasure. Cacioppo and Berntson (1994) demonstrated how much attitude research has suffered from a premature conceptualisation of attitudes along a single bipolar dimension. In contrast, affect research has a long tradition of conceptualising and assessing pleasure and displeasure separately (Nowlis & Nowlis, 1956). There exists no compelling evidence to break with this tradition at a time when attitude researchers are increasingly assessing attitudes along separate dimensions for positive and negative evaluations (Cacioppo & Berntson, 1994; Priester & Petty, 1996).

Manuscript received 12 March 2002

Revised manuscript received 18 June 2004

## REFERENCES

- Beebe-Center, J. G. (1932). *Psychology of pleasantness and unpleasantness*. New York: Van Nostrand.
- Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, *115*, 401–423.

- Diener, E., & Iran-Nejad, A. (1986). The relationship in experience between different types of affect. *Journal of Personality and Social Psychology, 50*, 1031–1038.
- Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., et al. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology, 53*, 712–717.
- Gannon, K. M., & Ostrom, T. M. (1996). How meaning is given to rating scales: The effects of response language on category activation. *Journal of Experimental Social Psychology, 32*, 337–360.
- Ito, T., Cacioppo, J. T., & Lang, P. J. (1998). Eliciting affect using the international affective picture system: Trajectories through evaluative space. *Personality and Social Psychology Bulletin, 24*, 855–879.
- Kaplan, K. J. (1972). On the ambivalence-indifference problem in attitude theory and measurement. *Psychological Bulletin, 77*, 361–372.
- Larsen, J. T., McGraw, A. P., & Cacioppo, J. T. (2001). Can people feel happy and sad at the same time? *Journal of Personality and Social Psychology, 81*, 684–696.
- Larsen, J. T., McGraw, A. P., Mellers, B. A., & Cacioppo, J. T. (2004). The agony of victory and thrill of defeat. Mixed emotional reactions to disappointing wins and relieving losses. *Psychological Science, 15*, 325–330.
- Meddis, R. (1972). Bipolar factors in affect adjective checklists. *British Journal of Social and Clinical Psychology, 11*, 178–184.
- Moskowitz, D. S., Pinard, G., Zuroff, D. C., Annable, L., & Young, S. N. (2001). The effect of tryptophan on social interaction in everyday life: A placebo-controlled study. *Neuropsychopharmacology, 25*, 277–289.
- Nowlis, V., & Nowlis, H. H. (1956). Toward an adequate taxonomy of personality attributes. *Annals of the New York Academy of Science, 65*, 345–355.
- Priester, J. R., & Petty, R. E. (1996). The gradual threshold model of ambivalence: Relating the positive and negative bases of attitudes to subjective ambivalence. *Journal of Personality & Social Psychology, 71*, 431–449.
- Raudenbush, S., Bryk, A., Cheong, Y. F., Congdon, R. (2000). *HLM 5: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.
- Reisenzein, R. (1992). A structuralist reconstruction of Wundt's three-dimensional theory of emotions. In H. Westmeyer (Ed.), *The structuralist program in psychology: Foundations and applications* (pp. 141–189). Toronto: Hogrefe.
- Reisenzein, R. (1995). On Oatley and Johnson-Laird's theory of emotion and hierarchical structures in the affective lexicon. *Cognition and Emotion, 9*, 383–416.
- Rozin, P. (1999). Preadaptation and the puzzles and properties of pleasure. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 109–133). New York: Russell Sage Foundation.
- Russell, J. A., & Carroll, J. M. (1999). On the bipolarity of positive and negative affect. *Psychological Bulletin, 125*, 3–30.
- Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect Grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology, 57*, 493–502.
- Schimmack, U. (1997). Das Berliner-Alltagssprachliche-Stimmungsinventar (BASTI): Ein Vorschlag zur kontextvaliden Erfassung von Stimmungen [The Everyday Language Mood Inventory (ELMI): Toward a content valid assessment of moods]. *Diagnostica, 43*, 150–173.
- Schimmack, U. (2001). Pleasure, displeasure, and mixed feelings? Are semantic opposites mutually exclusive? *Cognition and Emotion, 15*, 81–97.
- Schimmack, U. (in press). Methodological issues in the assessment of the affective component of subjective well being. In A. Ohn & M. van Dulmen (Eds.), *Handbook of methods in positive psychology*.
- Schimmack, U., Bockenholt, U., & Reisenzein, R. (2002). Response styles in affect ratings: Making a mountain out of a molehill. *Journal of Personality Assessment, 78*, 461–483.

- Schimmack, U., & Diener, E. (1997). Affect intensity: Separating intensity and frequency in repeatedly measured affect. *Journal of Personality and Social Psychology, 73*, 1313–1329.
- Schimmack, U., & Hartmann, K. (1997). Interindividual differences in the memory representation of emotions: Exploring the cognitive processes in repression. *Journal of Personality and Social Psychology, 73*, 1064–1079.
- Shapiro, D., Jamner, L. D., Goldstein, I. B., & Delfino, R. J. (2001). Striking a chord: Moods, blood pressure, and heart rate in everyday life. *Psychophysiology, 38*, 197–204.

