

Plain Texts as an Online Mood-Induction Procedure

Christopher Verheyen¹ and Anja S. Göritz²

¹Department of Organizational and Social Psychology, University of Erlangen-Nürnberg, Germany,

²Department of Work, Industrial & Organizational Psychology, University of Würzburg, Germany

Abstract. This experiment examined six short texts for their effectiveness in inducing positive or negative mood in an online environment. Compared to two control groups, the texts successfully changed mood in the desired direction for mood measured on a single-item self-report scale. In accordance with previous research, negative mood was induced more effectively than positive mood. In addition to the self-report mood measure, two non-self-report mood measures are used; namely, word-pleasantness ratings and subjective probabilities of negative life-events. Pleasantness ratings of uncommon words seemed to be a suitable measure for mood, whereas subjective probabilities of negative life-events were not.

Keywords: mood induction, online experiment, non-self-report mood measure, word-pleasantness ratings, subjective probabilities of negative life events

Introduction

To evaluate the impact of mood on people's behavior and thoughts in social cognition research, reliable procedures to induce the desired mood states are required. In the past, many such mood-induction procedures (MIPs) have been developed, most of which are described by Westermann, Spies, Stahl, and Hesse (1996) and by Otto (2000).

However, these MIPs have been developed and tested for use in face-to-face experiments only. The purpose of this study was to test whether mood induction is also feasible via the Web. If this was found to be the case, henceforward, Web-based experiments that examine the effect of mood as an independent variable would become possible. Because of the growing attractiveness of the Internet as a platform for conducting experiments, it is important to investigate whether established MIPs work equally well in an online setting, as well as to develop new MIPs for use via the Web.

Web experiments are an interesting methodology because they combine the advantages of the Internet with those of experiments. A Web experiment features all the advantages of computer-aided questionnaires, such as automated data handling, questionnaire routing, and real-time input validation. The number of participants has little impact on the costs of the study because no laboratory rooms and little personnel are needed. Furthermore, the Internet brings the experiments to the participants, and not vice versa: Respondents take part around the clock and in their familiar environment. Therefore, compared to traditional lab experiments with a reliance on sophomore samples, Web experiments typically draw larger and more heterogeneous

samples. However, Web experiments also entail disadvantages. The most prominent downside is that the spatial and temporal distance between investigator and participant does not allow the same amount of control over the participants and their surroundings as in the laboratory (Birnbaum, 2004b; Reips, 2001, 2002).

Instead of a leap-of-faith assumption that MIPs work comparably well online and offline, empirical evidence must establish this assumption as fact. While the availability of heterogeneous samples has rendered Web experimentation popular, the differences between online and offline experiments cautions one against the blind adoption of lab MIPs for use via the Web. In addition, in Web studies there is less control over the participants and their environment, causing smaller effects when compared to similar experiments conducted in a lab. In general, technical and environmental variance is higher in Web-based studies because the hardware and software of the participants differs widely, as do their Internet connection bandwidths and their social and physical surroundings (Birnbaum, 2004a, 2004b; Gosling, Vazire, Srivastava, & John, 2004). This increased error-variance in Web research suggests that online MIPs might be less effective than lab MIPs. On the other hand, online participation usually takes place in familiar environments of participants rather than in an unfamiliar lab so that demand effects might be lower and, thus, research results less prone to error.

Whereas many reliable MIPs have been developed for offline use, few have been developed or modified for online use (Göritz & Moser, 2006; Göritz, 2007). The need for reliable online MIPs is, therefore, obvious. Recently, Göritz and Moser (2006) investigated the effectiveness of the Velten-method (Velten, 1968), autobiographical recall, and

mood-suggestive photographs. Moreover, Göritz (2007) retested the Velten-method and photographs, and in addition examined cartoons, jokes, associations to emotion-laden words, and picture-illustrated emotive texts. In this study, several short texts were evaluated for their effectiveness in inducing either a positive or a negative mood. In general, plain texts not including other stimuli (e.g., pictures) have rarely been used, even as an offline MIP (Gerrards-Hesse, Spies, & Hesse, 1994; Westermann et al., 1996). To our knowledge, there is no previous online experiment in which plain texts have been used to alter mood. Göritz (2007, Study 4) used texts as an MIP, but they were accompanied by pictures.

Demands on an Online MIP

The first demand on an online MIP is, of course, that it is effective. The procedure must reliably produce the desired mood. Ideally, the MIP should be equally effective in inducing positive and negative mood states. Previous research has shown that it is easier to induce a negative mood state than a positive one (Gerrards-Hesse et al., 1994; Göritz & Moser, 2006; Westermann et al., 1996). The second demand is that the MIP is economical. Since Web experiments are cost-effective in general, financial economy is inherent. Furthermore, the target mood should be produced as quickly as possible. Brevity is desirable because in a psychological experiment the MIP often precedes the actual experiment and should not take time from it. Third, the MIP should not contain elements that prevent potential participants from taking part because their computer hardware or software or their connection to the Internet does not support the experiment (O'Neil & Penrod, 2001). Fourth, there should be no difference in dropout rates between the negative and the positive mood induction (Göritz, 2002; Göritz & Moser, 2006). For example, bias might be introduced into the experiment if the technique in use is considered ethically objectionable, resulting in participants becoming shocked by the stimuli and abandoning the study prematurely. Furthermore, a truly effective negative MIP might appear ineffective if participants quit the experiment prematurely because of their successfully induced negative mood. Systematic distortions like this should be avoided in every psychological experiment, but they are more likely to occur during a negative MIP that is administered in an online environment.

Finally, in each MIP, experimental demand should be minimal. Orne (1962) defines demand effects as the sum of clues that give away the experimental hypothesis to the participant and, thus, potentially influence his or her behavior. The impact of demand effects on the results of an MIP have been examined in detail for the Velten-method (Keane, 1986; Larsen & Sinnett, 1991). Larsen and Sinnett concluded that, for the Velten MIP, demand conditions somewhat inflated an already significant mood-induction effect. However, it cannot be ruled out that in MIPs that do

not use any precaution against demand effects, the measured effect is largely or solely caused by experimental demand and not by a real mood change. Therefore, demand characteristics should be minimized especially if a new MIP is tested in an online and, hence, rather unexplored environment.

Text-Based Stories for Online Mood Induction

Text-based stories as mood-inducing stimuli were chosen as the object of investigation because the meta-analyses of Westermann et al. (1996) and Gerrards-Hesse et al. (1994) revealed that MIPs in the category film/story are generally effective. In addition, an online MIP that is based on plain text does not exclude people for reasons of technical incompatibility. Also, the loading time of a Web page containing plain text is expected to be short and, as a result, compliance should not be impaired.

These advantages distinguish text-based stories from films, which have been used more often in the offline world and are even more promising with regard to effectiveness. Although a film MIP might seem more modern and appropriate in light of today's broadband Internet connections, texts feature some advantages that films cannot (yet) make up for. For example, a film MIP used online puts much higher demands on the hardware, software, and bandwidth of the server, as well as the client computer. Furthermore, the demands on the participant's computer are more specific. Computers without the appropriate media player software are excluded, as are computers without a soundcard or speaker – a configuration not uncommon in professional environments. Watching a film is also more intrusive for the environment of the participant. It is conceivable that many persons at their workplace would not take part in an online experiment if this study required watching and listening to a film of several minutes' length.

With regard to a desired minimization of experimental demand, text-based stories appear to be suitable as well. The real purpose of a mood-inducing story can be masked by a cover story quite easily. With the Velten MIP, a design using a deceptive cover story has been shown to be more resistant against experimental demand, especially if combined with non-self-report manipulation check measures (Larsen & Sinnett, 1991). It can be assumed that this is also the case with a text-based MIP.

Although Westermann et al. (1996) and Gerrards-Hesse et al. (1994) found that MIPs of the category film/story were among the most effective, only a few studies have actually used plain texts for inducing positive or negative mood states. Most studies included in the film/story category rely on films or on stories in the form of taped radio plays. All in all, only 13 mood inductions using text-based stories were found in the literature (Erber, 1991; Forgas, 1998; Johnson & Tversky, 1983; Knapp, 1986; Popp,

1988).¹ If the effect sizes (r) of 11 of the 13 aforementioned inductions are transformed into Fisher's z values and weighted by their sample size (Hedges & Olkin, 1985, pp. 230–231), a mean weighted effect of $r_m = .50$ is obtained.² Thereby, the effect size of the negative inductions is considerably higher ($r_m = .57$) than that of the positive inductions ($r_m = .35$). This is consistent with previous findings indicating that the creation of a positive mood is more difficult than the creation of a negative mood (Gerrards-Hesse et al., 1994; Göritz, 2007; Göritz & Moser, 2006; Westermann et al., 1996). However, the literature provides no exhaustive explanation for this phenomenon. A possible explanation could be mood maintenance tendencies (Handley, Lassiter, Nickell, & Herchenroeder, 2004) of the participants, leading to rather positive self-reported mood ratings in presumably neutral settings. Whatever the true reason(s) might be, we expect a positive mood induction to be more difficult than a negative one, especially in an online environment, because there is evidence that the baseline mood of online participants is more positive than that of offline participants taking part in a traditional lab experiment (Göritz, 2007).

The texts used in the 13 experiments differ considerably. Johnson and Tversky (1983) used a number of short texts in the form of newspaper articles with depressing content. Knapp (1986), Popp (1988), and Erber (1991) used three texts that were first used in the dissertation of Erber (1985). In these stories, the protagonist is a young female artist. In the positive text, a number of positive events are described that culminate in the protagonist receiving a scholarship to study art. The story designed to induce a negative mood describes how the young artist is overcome by a disabling illness in her freshman year in college. The neutral mood story simply describes how the person decides which college to attend. The experiment by Forgas (1998) differs from the others in that it was a field experiment, in which experimenters left envelopes on vacant library desks that contained a one-page text with either a humorous episode (positive mood induction), a sad narrative about a death from cancer (negative mood induction), or an information sheet about the library (control group). An informal review of the texts by people involved in this study resulted in the conviction that it should be possible to discover more effective texts.

The first goal of this study was to develop a reliable text-based online MIP that fulfills all requirements described earlier. This implies finding suitable texts that might be able to deteriorate or elate mood or – in the case of the control group – not influence mood at all. Furthermore, the texts must not be too long because shortness is a virtue in itself and reading a long text on the screen is as-

sumed to be more exhausting than reading it from printed paper. The second goal of this study was to evaluate two non-self-report measures in terms of their suitability to indicate different mood states. Besides using a cover-story, non-self-report measures are essential for reducing experimental demand in an MIP.

The hypotheses are as follows:

- Hypothesis 1: Mood after the induction procedure differs between the experimental condition groups in that mood in the groups that read a positive text is better than mood in the control groups, and mood in the control groups is better than mood in the groups that read a negative text. Mood states are operationalized by one self-report measure (rating scale) and two non-self-report measures (word pleasantness ratings and subjective probabilities of negative life events).
- Hypothesis 2: The induction of negative mood is more effective than the induction of positive mood.

Method

A between-subjects design with eight conditions including two control groups was chosen for the experiment. Groups 1 to 4 were positive mood induction groups, while Groups 5 and 6 underwent a negative induction. No other texts were tested in the experiment. This asymmetrical design was chosen because we expected that inducing a positive mood would be more difficult than inducing a negative one. Therefore, more effort was put into identifying and examining positive texts than negative ones. All positive and negative texts had been tested in multiple pretests and found to be promising. One of the control groups was what is called an untreated control group (UTCG), which means that people in that group did not receive any text to read, but instead explored and rated an external website. The other control group was a treated control group (NEUT). People in that group read a story in the same manner as those in the negative and positive induction groups, except that their story was assumed to be mood-neutral. The text used in NEUT was a slightly modified excerpt from the first chapter of Hawking (1997), describing humankind's picture of the universe from ancient times to the 17th century (699 words/4,674 characters including blank spaces). The first positive text (POS1) was a humorous, yet philosophical episode about a university professor who teaches his students about the important things in life (357/2,112). The second positive story (POS2) was about the relationship between a boy and his parents. The father speaks about the everyday challenges of the parents with their child in a

¹ Erber (1991), Johnson & Tversky (1983), and Popp (1988) conducted two mood inductions each, while Knapp (1986) did three and Forgas (1998) did four – hence the number of 13 identified mood inductions using text-based stories in only five articles. A detailed listing of all 13 studies including sample and effect sizes is available from the authors.

² For the two mood inductions of Erber (1991), a correlation coefficient cannot be calculated because of incomplete reporting. However, the negative induction was more successful than the positive one.

warmhearted and humorous manner, and he concludes that sometimes a single joyful event compensates for all troubles (698/4,084). Another positive text (POS3) consisted of 31 “howlers” from essays of elementary school students (577/3,559). The last positive text (POS4) was a story about a restaurant manager who always thinks positively. Even after an armed raid in which he becomes severely injured, he does not lose his optimism and humor (713/4,444). The first negative story (NEG1) was a first-person narrative by a woman whose father is suffering from Alzheimer’s disease. During a visit to her parents, the woman witnesses the great strains that her father’s disease puts on her mother (835/4,863). The second negative text (NEG2) described the unsavory aspects of the death penalty in the USA by depicting two executions in the electric chair (658/4,552).³

To keep experimental demand as low as possible, the effectiveness of the negative and positive stories in changing mood was assessed by two non-self-report measures labeled *word-pleasantness ratings* and *subjective probabilities of negative events*. Word-pleasantness ratings were based on Isen (1983, as cited in Isen, Johnson, Mertz, & Robinson, 1985) and on Forgas’ (1995, 2002) affect infusion model (AIM). Isen reported that, compared to a control group, the ratings of unfamiliar words were significantly more positive among participants in a positive mood. Forgas’ AIM postulates that a person’s mood is reflected in his or her evaluations of a stimulus and that the influence of mood on evaluative judgments is stronger the more unfamiliar the stimulus is. Based on these findings, we suspected that the more positive (negative) the mood of a person, the more (less) he or she would like an unfamiliar word. Moreover, we also suspected that the relationship between a person’s mood and his or her word rating would be stronger if the word is less familiar.

Subjective probabilities, our second non-self-report measure, were pilot tested by Johnson and Tversky (1983), who established that negative mood results in higher estimates of the frequency of many risks and other undesirable events. Accordingly, positive mood leads to a decrease in the judged frequency of risks. These findings have been endorsed by Wright and Bower (1992), who found happy (sad) people to report higher (lower) probabilities for positive events and lower (higher) probabilities for negative events. We, therefore, suspected that frequency estimates of negative life-events could be used to measure mood without creating experimental demand. Compared to people in a neutral mood, people in a negative (positive) mood are expected to estimate negative events to occur more (less) frequently.

Because the validity of the two non-self-report measures (i.e., word pleasantness ratings and subjective probabilities) is yet to be determined – especially with regard

to their use on the Internet – a self-report mood-measure was used. Because lengthy mood scales (e.g., Hampel, 1977) might increase demand effects, we opted for a single question on the participant’s current mood.

Participants were invited to participate by e-mail notifications that contained a personalized link to the study. The experiment’s first page was the introductory page, which estimated total participation time to be between 10 to 15 min. To reduce demand characteristics, a cover story was used (Larsen & Sinnett, 1991). The stated goal of the study was to find out the impact of creativity and context of an advertisement on the perception of advertisements. It was also stated that in the course of the experiment, people were expected to read a text and, thereafter, rate it and that this procedure would reveal the type of creative personality the participant could be classified as. On the second page, people were asked their gender, year of birth, and highest school degree obtained. After that, the participants in the untreated control group (UTCG) were asked to rate their current mood on a 7-point bipolar scale with the anchors 1 (*very negative*), 4 (*neutral*), and 7 (*very positive*). This scale was identical to the one presented to all participants after their respective mood inductions. Next came the actual MIP in the form of a text spread over two pages. People in the UTCG did not get to read a text, but instead were redirected to an external website that they were asked to explore for 3 min. When 3 min were up, people were automatically redirected to the experiment. After the texts or the website exploration, respectively, participants evaluated their current mood on the 7-point bipolar scale described above. For the purpose of perpetuating the credibility of the cover story and to further reduce demand effects that might have been created by this question, the relevance of the mood question was veiled by asking participants (on the same page) to rate the narrative quality of the text or – in the UTCG – the lucidity of the website. Next was a page with word-pleasantness ratings. One familiar word (year) and two increasingly unfamiliar words (terracotta, saté) were displayed. The frequencies of the words were extracted from Projekt Deutscher Wortschatz (2002).⁴ We assumed that frequency and familiarity of a word were directly linked. People indicated on a 7-point bipolar scale ranging from 1 (*very negative*) to 4 (*neutral*) to 7 (*very positive*) how the words affected them. Thereafter, participants made their probability ratings. The rating task was introduced by stating that, according to the Federal Statistical Office, 6962 persons were killed on German streets in 2001. Subsequently, participants were asked to estimate how many people in Germany were affected by several negative events. The negative events listed were divorce, sexual impotence, depression, heart attacks, unemployment,

³ The exact wording of all used texts is available from the authors.

⁴ The first uncommon word “Saté” (saté) belongs to frequency class 20, meaning that the most common German word “der” (the) is 2²⁰ times more common. With class 18, the word “Terracotta” (terracotta) is somewhat uncommon (class 18), while the word “Jahr” (year) is quite common (class 5).

Table 1. Means of self-reported and non-self-reported mood, number of invalid datasets, distress, length of stories, and processing time of induction for all conditions

	POS1	POS2	POS3	POS4	NEG1	NEG2	NEUT	UTCG
<i>n</i>	104	94	88	105	99	102	85	99
Self-reported postinduction mood (<i>SD</i>)	5.25 (1.30)	4.84 (1.51)	5.50 (1.33)	5.10 (1.33)	3.25 (1.31)	2.94 (1.50)	4.58 (1.41)	4.47 (1.20)
Pleasantness rating of <i>Saté</i> (<i>SD</i>)	4.49 (1.23)	4.64 (1.27)	4.59 (1.28)	4.51 (1.11)	4.21 (1.21)	4.21 1.18	4.47 (1.26)	4.40 (1.16)
Invalid datasets	85 45.0%	70 42.7%	67 43.2%	63 37.5%	87 46.8%	70 40.7%	103 54.8%	83 45.6%
Distress (<i>SD</i>)	1.53 (.70)	1.94 (.92)	1.68 (.85)	1.54 (.60)	1.74 (.82)	2.31 (.99)	2.37 (1.02)	2.34 (.93)
Length of story in characters (words)	2,112 (357)	4,084 (698)	3,559 (577)	4,444 (713)	4,863 (835)	4,552 (658)	4,674 (699)	^a
Processing time of induction (<i>SD</i>) [s]	107.39 (40.94)	188.86 (60.83)	194.99 (79.00)	196.40 (56.53)	193.68 (58.76)	199.35 (73.77)	181.59 (69.04)	190.26 (42.74)

Note. ^aThere was no story to read.

street crime, and accidental falls. These risks, or life problems, were chosen on account of their suitability to reflect positive as well as negative mood (Johnson & Tversky, 1983, Figure 3). To find out how much strain people experienced while reading the text on the screen, a page was designed for people to rate six statements on a unipolar scale ranging from 1 (*do not agree at all*) to 5 (*fully agree*). Five of the six items were introduced to compare the degree of distress and effort that the reading of the texts had caused. The sixth item measured experience in on-screen reading as a control variable, and was not present in the UTCG.

The subsequent page consisted of six humorous cartoons and instructions to mark the most appealing one. The cartoons' purpose was to remedy the possible effects of the negative mood induction, for ethical reasons. However, for the sake of experimental simplicity and as a reward for all participants who persevered up to this point, the cartoons were presented to all groups. After the cartoons, all participants were thoroughly debriefed and asked if they would like to receive a result summary. The last page thanked all participants for their cooperation.

Participants

Invitations were e-mailed to 8073 people from all walks of life.⁵ All persons had declared in an earlier telephone interview that they were interested in taking part in online surveys. The number of persons who visited, at minimum, the first page of the experiment was 1404 (17.4%). Of those, 1088 (77.5% of all first page visits) completed the study. Before any analyses were carried out, the data file was sift-

ed. A dataset was considered invalid if people dropped out before completing the on-screen reading distress page, if they missed more than four items combined on the relevant pages of the study, if their processing time for a single page was higher than 15 min, or if their processing time for the whole experiment was unrealistically low or high. After this procedure, 776 participants remained, of whom 282 (36.3%) were female. The average age was 38.3 years (*SD* = 12.05).

To test for systematic noncompliance, the proportion of invalid datasets in the four positive mood induction groups was compared to that in the two negative mood induction groups (cf. Table 1). The difference was not significant, $\chi^2(1, n = 1034) = .28, p = .60$.

Results

There were no significant differences in gender between the eight conditions, $\chi^2(7, n = 776) = 2.79, p = .90$. The same was true for age, $F(7, 763) = 1.42, p = .19$, and level of education, $\chi^2(7, n = 768) = 4.62, p = .70$.⁶ With regard to experience in on-screen reading of lengthy texts, there was also no significant difference between the seven text-groups, as a one-way analysis of variance (ANOVA) confirmed, $F(6, 669) = .90, p = .49$.

Effectiveness According to Self-Report

On account of the between-subjects design used, the effectiveness of the text-based mood induction was assessed by

⁵ The e-mail addresses were kindly provided by ForschungsWerk GmbH, Nürnberg, Germany (<http://www.forschungswerk.de>).

⁶ Eight participants in the category *no graduation (yet)* were not included in the Kruskal-Wallis test because of their low number and the ambiguous definition of the category.

Table 2. Effect sizes r of the positive and negative texts in relation to the control groups when measured with self-report and word pleasantness ratings

	Self-reported mood		Word pleasantness rating of <i>Saté</i>	
	NEUT	UTCG	NEUT	UTCG
POS1	.24	.30	.01	.04
POS2	.09	.13	.07	.10
POS3	.32	.38	.05	.08
POS4	.18	.24	.02	.05
NEG1	.44	.44	.10	.08
NEG2	.49	.49	.11	.08
Overall	.30	.34	.06	.07

comparing the self-report post-induction mood (rating scale) among the eight experimental condition groups. All positive and negative stories yielded a mood change in the desired direction when compared to the two control groups (Table 1, row 3). To test if these mood differences were statistically significant, we conducted a planned linear contrasts analysis (Rosenthal, Rosnow, & Rubin, 2000) with the following contrast coefficients: +.25, +.25, +.25, +.25, -.50, -.50, 0, 0 for POS1, POS2, POS3, POS4, NEG1, NEG2, NEUT, UTCG, respectively. The contrast was significant, $t(768) = 17.53, p < .001$. Therefore, our hypothesis that postinduction mood in the groups with the positive texts would be better than postinduction mood in the control groups, which in turn would be better than postinduction mood in the groups with the negative texts, was confirmed. To determine the magnitude of the mood change caused by each text, we calculated two effect sizes r for each text, one using NEUT and one using UTCG as a reference (Table 2, columns 2–3). According to the self-report measure, POS3 was the most effective positive story, followed by POS1, POS4, and POS2, respectively. Of the negative stories, NEG2 was slightly more effective than NEG1.

In order to compute an average weighted effect size r_m for all six mood inductions, we transformed each effect size r of the self-reported postinduction mood into a Fisher's z value and weighted it by its underlying sample size (Hedges & Olkin, 1985, pp. 230–231). This operation yielded an average weighted effect size of $r_m = .34$ when the UTCG was used as a reference. Using the same operation, but comparing each effect size r to the NEUT resulted in $r_m = .30$. When only the positive inductions were considered, the average effect size was $r_m = .26$ with the UTCG as a reference and $r_m = .21$ with the NEUT as a reference. An examination of the negative inductions only yielded $r_m = .47$ for both UTCG and NEUT as a reference. The difference between the mean effect sizes of the positive and the negative inductions was significant, whether NEUT ($z = 3.40, p < .001$) or UTCG ($z = 2.79, p < .01$) was used as a reference. Our hypothesis that the induction of a negative mood is more effective than the induction of a positive mood was confirmed.

Effectiveness According to Non-Self-Reports

The first non-self-report mood measure consisted of several word-pleasantness ratings. To analyze whether differences of the ratings between the experimental groups were in accordance with our hypothesis, we conducted a planned contrasts analysis with the same contrast coefficients as in the analysis of the self-report mood measure. There was no significant difference for the common word *Jahr* (year), $t(757) = .25, p = .81$, which was in line with our hypothesis because only ratings of uncommon words were expected to be susceptible to mood. The first somewhat uncommon word *Terracotta* yielded no significant effect, either, $t(764) = .24, p = .81$. However, the contrast was significant for the very uncommon word *Saté*, $t(768) = 3.32, p = .001$. The similarity of the ratings of *Saté* to the self-reported mood is depicted in Figure 1.

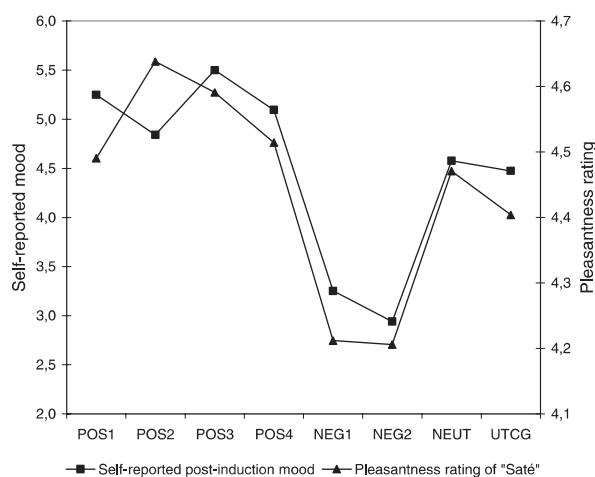


Figure 1. Comparison of self-reported post-induction mood and pleasantness rating of the uncommon word *Saté*.

To analyze the second non-self-report mood measure – subjective probabilities – each estimation was first logarithmized because of the wide spread of the participants' estimations (Johnson & Tversky, 1983). Then, the means of the logarithmized values were computed for each participant. A planned contrasts analysis with the same contrast coefficients used in the prior analyses yielded no effect, $t(761) = 1.26, p = .21$.

Discussion

According to the self-reported mood measure, all stories were successful in inducing the desired mood, no matter if they were compared to the treated control group with the neutral text (NEUT) or to the untreated control group that explored an external website (UTCG). POS3 was the most

effective positive story, followed by POS1, POS4, and POS2. However, if a single story is to be recommended, it would be POS1. This story is almost equally effective as POS3, but it is considerably shorter and it causes the least distress (Table 1). With the negative stories, NEG2 is more effective than NEG1. However, the advantage of NEG2 might come at a cost because it causes significantly more distress than NEG1, $t(199) = -4.43, p < .001$. In the future, it would be worth investigating whether the higher distress caused by NEG2 is the result of a difference in the type and number of influenced emotions. NEG2 is a text about the death penalty in the USA, whereas NEG1 is a first-person narrative of a woman whose father is suffering from Alzheimer's disease. However, this study did not attempt to answer this.

In self-reported mood in reference to the UTCG, the positive inductions are considerably less effective, with $r_m = .26$, than the negative inductions, with $r_m = .47$. This corresponds to previous findings with both offline (Gerrards-Hesse et al., 1994; Kenealy, 1986; Larsen & Sinnett, 1991; Westermann et al., 1996) and online MIPs (Göritz, 2007; Göritz & Moser, 2006) MIPs.

If the online MIPs tested in Göritz and Moser (2006) are compared to the text-based MIP developed in this study, the latter is more promising. The positive Velten-induction in Göritz and Moser (2006) was less effective ($r = .11$) than the positive texts, as were photographs and autobiographical recall ($r = .12$). The same is true for the negative inductions, where Velten ($r = .28$), photographs ($r = .41$), and autobiographical recall were also less effective. However, this comparison is made with caution because the effects in Göritz and Moser (2006) were based on mood measurement on a standardized scale, while the ones in this study were based on a single-item mood measure.

If the mean effect size of all six mood inductions in the UTCG ($r_m = .34$) is compared to the mean effect size ($r_m = .50$) of 11 of the 13 offline studies that previously used plain texts for mood induction (see "Text-Based Stories for Online Mood Induction" above), the mood manipulation in this study was significantly less effective ($z = 3.54, p < .001$). At least in part, this difference might be caused by publication bias. In fact, none of the identified 13 studies using texts for offline mood induction was published as a methodological article examining an MIP's effectiveness. Instead, they all investigated the influence of mood on other dependent variables. For these types of studies, publication bias can be enormous because most journals do not publish papers on the impact of mood on other variables if the mood induction did not work (Westermann et al., 1996). A reanalysis of Westermann et al.'s data by Göritz and Moser (2006, p. 894, footnote 7) yielded a considerable difference in effect sizes between methodological papers ($r_m = .37$) and other articles ($r_m = .48$). In fact, the mean effect size of $r_m = .37$ for methodological papers roughly corresponds to this study's mean effect size of $r_m = .34$.

The effect sizes of the mood inductions in this experiment can also be compared to Göritz (2007), who used some of the

texts also used in this study but on another sample. The positive induction in Göritz (2007, Study 4) used the same text that was used in POS3, albeit accompanied by two cartoons featuring children. The negative induction used the same text as in NEG2, but illustrated by two photographs depicting an execution. The neutral induction used the text of NEUT, again supplemented by two illustrative pictures. Because Göritz (2007) also used the same single-item self-report scale that was used here, we can directly compare the effect sizes. The positive induction in Göritz's study yielded a smaller effect ($r = .12$) than in this study's POS3 ($r = .32$). Along the same vein, the negative induction yielded a smaller effect ($r = .10$) than in this study's NEG2 ($r = .49$). The reasons for this discrepancy remain unclear. It is hardly conceivable that the addition of two pictures to each text alone diminished effectiveness, especially because the pictures were intended to intensify effectiveness. Perhaps, participants were more engaged in this study because, in contrast to Göritz (2007, Study 4), the alleged purpose of this study was to assess their creativity.

The study indicated that the positive induction ($r_m = .26$, in the UTCG) was significantly less effective than the negative induction ($r_m = .47$). This confirms the findings of previous offline studies using texts as an MIP, as well as those of other offline MIPs (Gerrards-Hesse et al., 1994; Westermann et al., 1996). The three MIPs (Velten, photographs, and autobiographical recall) also showed this difference when used online (Göritz & Moser, 2006). Göritz (2007) has suggested that this might, in part, result from the better baseline mood of online participants compared to offline participants. Göritz found evidence that the baseline mood of participants in online studies is generally more positive than that of participants in the laboratory. The preinduction mood of 4.71 found in this study's UTCG points in the same direction. A baseline mood of 4.71 is higher than the 4.00 that was anchored *neutral* in the scale. This might either be caused by participants' prior surfing on the Web, or by a selection effect, in the sense that people in a positive mood are more likely to take part in online studies. Surfing the Web for leisure might have an effect similar to a positive MIP because the Internet is commonly used for engaging in several types of rewarding behavior. An additional reason for the better baseline mood of online participants could be that taking part in online studies might, on average, be more pleasant than taking part in offline studies because it is often more voluntary and less tedious. In contrast, offline participants from "sophomore samples" could be in worse moods because they frequently take part for course credit, and they need to spend more time physically getting to the lab. If further studies prove that online participants are, indeed, generally in a more positive baseline mood than offline participants, the Internet could be of limited use for positive mood manipulations because of the ceiling effect that would make it difficult to further improve their already positive mood.

In this experiment, we tested two different non-self-report measures for mood that have not been used in an online

environment before. We found no evidence for subjective probabilities of negative life events being a suitable measure of mood. The fact that Göritz (2007, Study 4) did not find proof for the suitability of a very similar measure leads to the conclusion that this measure is not a reliable mood indicator – at least when used in an online environment. It could also be the case that the mood of the participants had already returned to its original state at this point of the experiment. Mood effects are known to be rather fragile (Frost & Green, 1982; Isen & Gorgoglione, 1983; Isen, Clark & Schwartz, 1976). The effect of the induction could have faded away during the two tasks between the texts and the subjective probabilities page. An alternative account would be that subjective probabilities are indeed a valid measure for mood, and the mood effects reflected by the self-report item discussed above are pseudo-effects from experimental demand. However, we reject this idea for four reasons. First, effect sizes derived from self-reports differed among the different texts used (Table 2). If experimental demand alone was responsible for the effects found, it could be expected that the effect sizes would be homogeneous across the different texts used – which was not the case. Second, throughout the whole study we kept experimental demand low by not using a lengthy mood scale and by providing a cover story to mask the real purpose of the experiment. Third, the second non-self-report mood measure, word-pleasantness ratings for the uncommon word *Saté*, yielded a significant effect in line with our expectation. Figure 1 shows that the word ratings closely resemble the self-reported mood of our participants. Thus, a genuine mood change has taken place. The finding that the ratings of *Saté* obviously reflect postinduction mood while the ratings of the more common *Terracotta* do not indicates that a word has to be sufficiently uncommon to be a suitable measure of mood. Fourth, the conclusion that a genuine mood change was effectuated is supported by Göritz (2007, Study 4), who also found significant effects for mood self-reports as well as for pleasantness ratings of the word *Saté*, but no effect for subjective probabilities of life events.

Limitations

Self-reported mood in this study is based on a simple, single-item, 7-point bipolar scale with a neutral middle anchor. While this is often not the best practice to measure mood (see the literature on the ambivalence-indifference conflict, e.g., Feldmann Barrett & Russell, 1998; Russell & Carroll, 1999), we deliberately decided to use such a single-item-scale for two reasons. First, we wanted to keep the whole study as short as possible and, therefore, refrained from using a lengthy mood scale. Second, and more importantly, a lengthy scale would have very likely raised experimental-demand effects. For the same reason, we did not measure preinduction mood in the experimental groups, but only in the untreated control group. The weakness of this approach is that we have no information about

the within-subject change in mood. Therefore, it is possible, but in view of the random assignment of participants to conditions, unlikely, that the differences in postinduction mood are the result of a difference in preinduction mood. For future research, we recommend that mood-inducing texts also be evaluated in a within-subjects design using standardized mood scales; for example, the adjective scales for the assessment of mood (SES) by Hampel (1977).

Another potential weakness lies in the selection of our cover story. It is possible that the topic of “creativity” bears a positive mood bias and, therefore, impeded positive mood effects and facilitated negative mood effects. Even if we consider this bias to exist, its effects are at best mild. Still, we recommend developing and testing more neutral cover stories for further research.

Ethics

The purpose of the study was disclosed to the participants on the last Web page. In addition, several precautions were taken to minimize any harmful impact of the experiment on participants. First, the invitation e-mails emphasized the fact that taking part was voluntary and that the experiment, once begun, could be terminated any time. Second, the contents of the used texts were no more shocking or disgusting than a comparable newspaper article. This assumption was confirmed through personal feedback by the pretesters. Lastly, humorous cartoons were presented to every participant at the end of the experiment to remedy any adverse effects possibly caused by the mood induction.

Several facts confirm that the texts, especially those used to induce negative mood, did not cause any harm to the participants. First, no participant complained about the texts or their effect in any way. Second, there were no significant differences in compliance between the positive and the negative induction groups or between the negative and the neutral groups. Third, the mean on the distress scale was generally low, that is, 1.9 on a 5-point scale. The mean distress in NEG1 was within the range of the positive groups and below those of the control groups. Distress in NEG2 was comparable to the control groups. However, if more shocking material for a negative mood induction is to be used (e.g., explicit pictures or videos), participants should be warned beforehand and given the opportunity to quit the study or skip this part of the study (Göritz, 2007).

Summary

In this study, using a large and heterogeneous sample, different texts were developed and evaluated to be used for inducing positive as well as negative mood, via the Web. Through using both self-report and non-self-report measures of mood, it was established that the texts effect a genuine mood change. In future studies, text-based mood-induction procedures can probably be improved by optimiz-

ing texts already available as well as finding even more effective new texts.

Authors' Note

This article is based on the first author's final thesis, supervised by the second author, at the Department of Organizational and Social Psychology, University of Erlangen-Nürnberg, Germany. This work has been supported by a University of Erlangen-Nürnberg post-doctoral scholarship (HWP) to Göritz.

References

- Birnbaum, M.H. (2004a). Human research and data collection via the Internet. *Annual Review of Psychology*, *55*, 803–832.
- Birnbaum, M.H. (2004b). Methodological and ethical issues in conducting social psychology research via the Internet. In C. Sansone, C.C. Morf, & A.T. Panter (Eds.), *Handbook of methods in social psychology* (pp. 359–382). Thousand Oaks, CA: Sage.
- Erber, R. (1985). *Choosing among multiple categories: The effect of mood on category accessibility, inference, and affect*. Unpublished doctoral dissertation, Carnegie-Mellon University, Pittsburgh, PA.
- Erber, R. (1991). Affective and semantic priming: Effects of mood on category accessibility and inference. *Journal of Experimental Social Psychology*, *27*, 480–498.
- Feldman Barrett, L., & Russell, J.A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology*, *74*, 967–984.
- Forgas, J.P. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin*, *117*, 39–66.
- Forgas, J.P. (1998). Asking nicely? The effects of mood on responding to more or less polite requests. *Personality and Social Psychology Bulletin*, *24*, 173–185.
- Forgas, J.P. (2002). Feeling and doing: Affective influences on interpersonal behavior. *Psychological Inquiry*, *13*(1), 1–28.
- Frost, R.O., & Green, M.L. (1982). Velten mood-induction procedure effects: Duration and postexperimental removal. *Personality and Social Psychology Bulletin*, *8*, 341–347.
- Gerrards-Hesse, A., Spies, K., & Hesse, F.W. (1994). Experimental inductions of emotional states and their effectiveness: A review. *British Journal of Psychology*, *85*, 55–78.
- Göritz, A.S. (2002). Stimmungsinduktion über das WWW [Mood induction over the WWW]. *Report Psychologie*, *27*, 192–202.
- Göritz, A.S. (2007). The induction of mood via the WWW. *Motivation and Emotion*, *31*, 35–47.
- Göritz, A.S., & Moser, K. (2006). Web-based mood induction. *Cognition and Emotion*, *20*, 887–896.
- Gosling, S.D., Vazire, S., Srivastava, S., & John, O.P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, *59*, 93–104.
- Hampel, R. (1977). Adjektiv-Skalen zur Einschätzung der Stimmung (SES) [Adjective scales for the assessment of mood]. *Diagnostica*, *23*, 43–60.
- Handley, I.A., Lassiter, G.D., Nickell, E.F., & Herchenroeder, L.M. (2004). Affect and automatic mood maintenance. *Journal of Experimental Social Psychology*, *40*, 106–112.
- Hawking, S. (1997). *Die Illustrierte Kurze Geschichte der Zeit* [The illustrated a brief history of time] (Updated and expanded edition). Reinbek, Germany: Rowohlt.
- Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
- Isen, A.M. (1983). The influence of positive affect on cognitive association. Stanford conference on aptitude, learning, and instruction: Affective and cognitive processes, R.E. Snow, chair [as cited in Isen, Johnson, Mertz, & Robinson, 1985].
- Isen, A.M., Clark, M., & Schwartz, M.F. (1976). Duration of the effect of good mood on helping: "Footprints on the sands of time." *Journal of Personality and Social Psychology*, *34*, 385–393.
- Isen, A.M., & Gorgoglione, J.M. (1983). Some specific effects of four affect-induction procedures. *Personality and Social Psychology Bulletin*, *9*, 136–143.
- Isen A.M., Johnson, M.M.S., Mertz, E., & Robinson, G.F. (1985). The influence of positive affect on the unusualness of word associations. *Journal of Personality and Social Psychology*, *48*, 1413–1426.
- Johnson, E.J., & Tversky, A. (1983). Affect, generalization, and the perception of risk. *Journal of Personality and Social Psychology*, *45*, 20–31.
- Kenealy, P.M. (1986). The Velten mood-induction procedure: A methodological review. *Motivation and Emotion*, *10*, 315–335.
- Knapp, A. (1986). Die Auswirkungen emotionaler Zustände auf das Lösen eines sozialen Dilemmas [The effects of emotional states on the dissolving of a social dilemma]. *Zeitschrift für Sozialpsychologie*, *17*, 160–172.
- Larsen, R.J., & Sinnett, L.M. (1991). Meta-analysis of experimental manipulations: Some factors affecting the Velten mood-induction procedure. *Personality and Social Psychology Bulletin*, *17*, 323–334.
- O'Neil, K.M., & Penrod, S.D. (2001). Methodological variables in web-based research that may affect results: Sample type, monetary incentives, and personal information. *Behavior Research Methods, Instruments, and Computers*, *33*, 226–233.
- Orne, M.T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*, 776–783.
- Otto, J.H. (2000). Induktionsverfahren [Induction procedures]. In J.H. Otto, H.A. Euler, & H. Mandl (Eds.), *Emotionspsychologie. Ein Handbuch* [Emotional psychology: A handbook] (pp. 395–408). Weinheim, Germany: Psychologie Verlags Union.
- Popp, M. (1988). Wirkung von Informationsbedingung, emotionalem Zustand und externer Speichermöglichkeit auf das Lösen sozialer Fallen [Effect of information condition, emotional state, and possibility of external storage on the dissolving of social traps]. *Archiv für Psychologie*, *140*, 33–51.
- Projekt Deutscher Wortschatz. (2002). Retrieved August 6, 2002, from <http://wortschatz.informatik.uni-leipzig.de/index.html>
- Reips, U.-D. (2001). The Web Experimental Psychology Lab: Five years of data collection on the Internet. *Behavior Research Methods, Instruments, and Computers*, *33*, 201–211.
- Reips, U.-D. (2002). Standards for Internet-based experimenting. *Experimental Psychology*, *49*, 243–256.
- Rosenthal, R., Rosnow, R.L., & Rubin, D.B. (2000). *Contrasts*

- and effect sizes in behavioral research: A correlational approach.* Cambridge, UK: Cambridge University Press.
- Russell, J.A., & Carroll, J.M. (1999). On the bipolarity of positive and negative affect. *Psychological Bulletin*, *125*, 3–30.
- Velten, E. (1968). A laboratory task for induction of mood states. *Behavior Research and Therapy*, *6*, 473–482.
- Westermann, R., Spies, K., Stahl, G., & Hesse, F.W. (1996). Relative effectiveness and validity of mood induction procedures: A meta-analysis. *European Journal of Social Psychology*, *26*, 557–580.
- Wright, W.F., & Bower, G.H. (1992). Mood effects on subjective probability assessment. *Organizational Behavior and Human Decision Processes*, *52*, 276–291.

Received February 15, 2007
Final revision received July 11, 2008
Accepted July 14, 2008

Christopher Verheyen

Department of Organizational and Social Psychology
Universität Erlangen-Nürnberg
Lange Gasse 20
D-90403 Nürnberg
Germany
Tel. +49 911 530-2247
Fax +49 911 530-2243
E-mail christopher.verheyen@wiso.uni-erlangen.de