European Psychologist

The Effects of a Caffeine Placebo and Experimenter Expectation on Blood Pressure, Heart Rate, Well-Being, and Cognitive Performance

Harald Walach[1,2], Stefan Schmidt[1], Yvonne-Michelle Bihr[2], and Susanne Wiesch[2]
[1]*Department of Environmental Medicine and Hospital Epidemiology, University Hospital Freiburg, Germany,*
[2]*Department of Psychology, University of Freiburg, Germany*

We studied the effect of experimenter expectations and different instructions in a balanced placebo design. 157 subjects were randomized into a 2 × 4 factorial design. Two experimenters were led to expect placebos either to produce physiological effects or not (pro- vs. antiplacebo). All subjects except a control group received a caffeine placebo. They were either made to expect coffee, no coffee, or were in a double-blind condition. Dependent measures were blood pressure, heart rate, well-being, and a cognitive task. There was one main effect on the instruction factor (p = 0.03) with the group "told no caffeine" reporting significantly better well-being. There was one main effect on the experimenter factor with subjects instructed by experimenter "proplacebo" having higher systolic blood pressure (p = 0.008). There was one interaction with subjects instructed by experimenter "proplacebo" to receive coffee doing worse in the cognitive task than the rest. Subjects instructed by experimenter "antiplacebo" were significantly less likely to believe the experimental instruction, and that mostly if they had been instructed to receive coffee. Contrary to the literature we could not show an effect of instruction, but there was an effect of experimenters. It is likely, however, that these experimenter effects were not due to experimental manipulations, but to the difference in personalities.

Keywords: Caffeine, balanced placebo design, experimenter, instruction, expectation, placebo effect.

Expectations of substance-specific effects seem to trigger many physiological and psychological reactions (Kirsch, 1997). Moreover, these reactions seem to be quite independent of the substance given. This has been extensively studied in studies using the so-called balanced placebo design. In this design subjects receiving a test substance or a placebo are either informed that they are receiving this substance or placebo, or vice versa. Thereby, effects of expectations can be separated from actual drug effects. A review by Marlatt and Rohsenow (1980) suggests that there are specific effects of the expectation to receive alcohol consistent with the social stereotype. This result of a narrative review has been substantiated by a formal metaanalysis (Hull & Bond, 1986) summarizing the results of 36 studies. The authors found small but significant effects of alcohol consumption and expectation to receive alcohol (alcohol $g = 0.18$; $z = 5.1$; expectation $g = 0.08$; $z = 2.03$). Interestingly, no interactions occurred. Expectation showed a significant effect in antisocial behavior ($g = 0.4$), sexual arousal ($g = 0.3$),

Harald Walach holds a diploma degree in psychology from the University of Freiburg, Germany, and PhD degrees in clinical psychology from the University of Basel and in the theory and history of science from the University of Vienna. He completed his habilitation 1998 at the University of Freiburg. Since 1999 he is Director of Research and Evaluation of Complementary Medicine at the Institute of Environmental Medicine and Hospital Epidemiology, University Hospital Freiburg. His research interests include the evaluation of complementary medicine, especially nonspecific effects of therapy and placebo effects.

Correspondence concerning this article should be addressed to PD Dr. Dr. Harald Walach, University Hospital Freiburg, Department of Environmental Medicine and Hospital Epidemiology, Hugstetterstr. 55, D-79106 Freiburg, Germany (tel. +49 761 270-5497/-7223, fax +49 761 270-7224, E-mail walach@ukl.uni-freiburg.de).

European Psychologist, Vol. 6, No. 1, March 2001, pp. 15–25

and craving or consumption of more supposedly alcoholic drinks ($g = 0.5$), whereas the effects of the consumption of alcohol were negligible in these parameters though seen in cognitive processing ($g = 0.5$), aggressiveness ($g = 0.2$), motor ability ($g = 0.6$), and physical sensations ($g = 0.6$). In other words, alcohol shows its drug-dependent effects in areas in which it is known to affect brain functions. The expectation of receiving alcohol, on the other hand, is effective in areas where the social stereotype and folk wisdom suggests alcohol has its effects. One conclusion from this metaanalysis was that a fully balanced placebo design is not necessary to study expectancy effects, since there were no relevant and significant interactions. Instead, it would be sufficient to use only the placebo part and manipulate expectations.

Studies on the effects of expectancies have been also carried out in other areas as well, like expectation and consumption of cannabis (Camí et al., 1991), nicotine (Hughes et al., 1989; Kirsch & Rosadino, 1993), and caffeine (Fillmore & Vogel-Sprott, 1992; Fillmore et al., 1994; Kirsch & Weixel, 1988; Lienert, 1955; Murray, 1988; Zwyghuizen-Doorenbos et al., 1990). All of these studies demonstrated substantial effects of expectancies. The studies of Fillmore et al. (1992, 1994) used only one-half of the balanced placebo design: Besides a zero-control group with no experimental manipulation, all groups received placebo coffee. The first of these studies showed a significant effect of expectation on motor function and well-being. The second experiment studied the interaction of drug expectations and contradictory instructions. Two groups of subjects received a caffeine placebo and either the instruction, coffee would enhance or diminish their functions; two groups received the same instructions with an alcohol placebo; another group served as a control group, receiving no substance and no instruction. As expected, there was an interaction between type of placebo and instruction. Whereas subjects receiving caffeine placebo had clear-cut (better) results on a rotor pursuit task when instructed to expect enhancement of function, subjects on alcohol placebo did better when expecting a decrease of function. These effects were significant against the control group. Obviously, instructions consistent with the social stereotype, i. e., alcohol diminishes cognitive function, caffeine arouses and thereby enhances cognitive functions, produce these effects more easily as instructions geared toward the opposite direction.

Kirsch and Weixel (1988) used the same approach, but added a double-blind group, which also received a caffeine placebo with the instruction that they might randomly and blindly either receive caffeine or no caffeine. Apart from that they employed different dosages of caffeine placebos together with the expectation of caffeine. Five variables showed significant changes in the experimental groups compared with the control group that received no substance and no expectancy manipulation: heart rate, diastolic blood pressure, alertness, tension, and subjective probability to have received caffeine. Moreover, there were dose-dependent effects. While in the open-instruction group the effects were initially stronger with higher doses and then levelled out, in the double-blind condition effects were stronger with higher doses. This lead the authors to question the validity of the double-blind paradigm for evaluating drugs in clinical trials: "If double-blind administration produces psychological effects that in some instances are opposite to those produced by clinical administration of drugs … then double-blind procedures may not be appropriate methods by which to evaluate drug effects" (Kirsch and Weixel, 1988, p. 323). These effects seem to be quite stable even in other cultural contexts. Lienert, who used a balanced placebo design already in 1955, found that positive suggestions were as effective as caffeine itself in altering heart rate and self-reported general effects.

Taken together, these studies seem to suggest:

a. A fairly clearcut and consistent effect of the expectancy to receive coffee, even if no coffee is employed;
b. Differential effects of double-blind administration of a caffeine placebo;
c. Significant effects against zero controls in both psychological and physiological variables and in behavioral variables as well.

While the existence of expectancy effects in balanced placebo designs or parts thereof seem to be uncontested, apart from one exception (Kienle & Kiene, 1996) it is rather unclear how these effects are in fact brought about. We wanted to study the process by which expectancy effects are conveyed and modified. Relying on the results of the previously published studies, we designed an experiment with different caffeine instructions, together with a manipulation of expectancy in the experimenters involved in the study.

Method

Design

We used a 2 × 4 factorial design. The first factor represented the *expectation of the experimenters involved*, the

European Psychologist, Vol. 6, No. 1, March 2001, pp. 15–25

second factor the *instructions and experimental manipulations of the subjects*. We hired two students as experimenters who had no knowledge of the relevant literature and who were completely blind to the real purpose of the study. Both were males, of the same age, and comparable in their appearance at first sight. They were reimbursed for their work and instructed that the purpose of the study would be to study the placebo effect in a setting in which both coffee and a placebo was used to purportedly assess "effects of caffeine on working at the computer screen." They were located in separate rooms in different buildings, did not meet each other, but only knew that someone else was also carrying out the same experiment in parallel. They received at random two contradictory information:

Experimenter Expectancy (Factor 1)

a. *Proplacebo:* A short paper of scientific appearance was drafted containing the information, backed by references, that caffeine is shown to produce an alteration in physiological and cognitive functions and that caffeine placebos also produce effects in these directions. Moreover, it was emphasized that placebo effects exist and can be quite as large as the effects of the real substance.

b. *Antiplacebo:* A paper of about the same length was drafted saying that while the effects of caffeine on the physiological and cognitive system were clear-cut, the effects of placebos were probably due to artefacts and not real effects, backed by the appropriate literature (Kienle & Kiene, 1996; Kienle, 1996; Kienle, 1995).

Subject Expectancy (Factor 2)

Subjects were informed by leaflets and subject information flyers that the purpose of the study was to study the effects of caffeine on performance at the computer screen. They were reimbursed (DM 30 for 1.5 hours), and after having given written informed consent they were randomly allocated to one of four groups:

- *True information:* subjects received a caffeine placebo and were told so.
- *False information:* subjects received a caffeine placebo, but were told that they would receive coffee.
- *Double-blind group:* subjects received a caffeine placebo and were told that neither they nor the experimenter knew whether they would receive real coffee or decaffeinated coffee.
- *Zero-control:* no substance was given.

Measures

As dependent variables we measured both before and after experimental manipulation:

a. *Blood pressure* (systolic and diastolic) using an automatic measuring device ("boso-carat" by BOSO, Jungingen, Germany), which automatically inflates and deflates the cuff and displays the reading on a display. Two readings were taken three minutes apart from each other and averaged to yield blood-pressure scores.

b. *Heart rate*, using the same device, simultaneously with the blood pressure.

c. *General well-being*, using the Basler Befindlichkeitsskala (Hobi, 1985). This scale is virtually identical to that used by Kirsch and Weixel (1988). It is frequently used in the settings of psychopharmacological studies and is well known for its sensitivity to short-term changes. It measures general well-being in 16 pairs of adjectives arranged as a semantic differential. It measures the four dimensions alertness, extraversion, vitality, inner balance and can also be summed to one general dimension.

As secondary variables introduced to make the coverstory trustworthy we measured both before and after experimental manipulation:

d. *Performance on a self-devised test.* The test consisted in finding misprints in a text which was displayed on the screen for 4 minutes. For pre- and posttest, different sections of the same text with the same number of words were used containing the same amount and approximately the same structures of misprints. The measure was the number of misprints found in the time given.

e. *Performance on a computer game* (freeware, called "Wally the worm"). The measure was the number of points scored in the game.

Apart from these dependent variables, we also assessed some moderator variables before or after the experiment.

f. *General expectations* of the subjects about the general effects of coffee, before the experiment.

g. *Specific expectations* about the specific effects of the beverage taken (only experimental groups a–c) on all variables a–e, after the pretest and group allocation, but before consumption of the beverage on a 5-point, balanced Likert scale.

h. *Subjective probability* of caffeine intake after the posttest on a visual analog scale with the anchors "surely no caffeine" and "surely caffeine."
i. *Descriptive data:* demographic data, motivation in participation (money, curiosity, boredom scientific interest, or other).
j. *Inclusion- and exclusion criteria* (see below) in a questionnaire at the beginning of the experiment and as part of the procedure of seeking written informed consent.

Experimenters also gave data on their personality type (Freiburger Persönlichkeitsinventar, FPI, one of the most widely used scales in Germany by Fahrenberg et al., 1989) and about motivations, background, and expectations.

After the whole experiment was over, subjects were debriefed in a telephone interview. They were informed about the real purpose of the experiment and about the fact that no caffeine was used. Before debriefing they were asked about their expectations, the quality of interaction with the experimenter, their perception of the experimenter, the plausibility of the procedures and the overt experimental question, their own hypotheses about the experiment, and whether they had seen through the purpose of the experiment and the deception.

Procedures

Subjects were recruited on the campus of the university and in the city of Freiburg during early summer 1997. Interested subjects were contacted and informed about procedures and inclusion and exclusion criteria. Inclusion criteria were regular intake of coffee (at least one cup per day) and a minimum age of 16. Exclusion criteria were a psychiatric diagnosis, circulatory and/or heart problems, alcohol dependence or abuse, high blood pressure, serious and life-threatening diseases, intake of alcohol, antihistaminics, or psychotropic/illegal drugs on the day of the experiment, intake of caffeine during the last 4 hours before the experiment.

At no time were subjects informed about the true purpose of the experiment, about its design, or about the fact that only decaffeinated coffee was being used throughout. Since an experiment like this had never been done at our university before and the issue had not been brought up in teaching, it is very plausible to assume that all the subjects were blind as to the experimental procedures.

The two experimenters were trained thoroughly by the junior authors (SW and NB) in all experimental procedures. They were given their instructions in a circumstantial manner while being instructed about the experimental procedure and in writing to be taken home and studied. Each experimenter did some trial runs with pilot subjects to make sure that they had understood the procedure. Experimenters were blind as to the real purpose of the experiment and to the fact that only decaffeinated coffee was being used. The junior authors, who knew about the purpose and the design of the experiment, were instructed to reduce contact with the experimenters during the experiment to a minimum.

A random code was used to prepare 80 opaque envelopes for each experimenter, which contained the experimental instructions to the subjects, balanced for the four experimental conditions. The envelopes were numbered and opened sequentially.

Arriving subjects were greeted, questioned as to inclusion and exclusion criteria, and asked for written informed consent. Subjects then were first explained the experimental procedure. They were shown the blood-pressure-measuring apparatus in a dummy measurement, so as to make them acquainted and prevent artificial high initial readings. Then they filled in the questionnaires. After that, blood-pressure and heart-rate readings were taken twice within 3 minutes. Subjects were left to themselves and the computer screen pretests were run. Then the experimenter entered to open the coded envelope in front of the subject, reading the experimental assignment and brewing the "coffee" accordingly. After having provided the "coffee" to the subjects, given the consumption instructions, the experimenter left the subjects to themselves.

The coffee used was a widely used decaffeinated coffee from Costa Rica. In a pretrial run we had made sure that the placebo coffee resembled real coffee in taste, appearance, and smell. The dosage used was always three heaped spoonfuls of coffee. Coffee was brewed by a filter coffee-brewing machine. Subjects had 5 minutes to drink the coffee, black, without sugar; artificial sweetener was provided on demand. No coffee whitener or milk was allowed. After this drinking period subjects were required to wait for 5 more minutes, to "allow the caffeine to take effect." During this time they could read a travel magazine.

After this waiting period, the series of posttests followed in the same order as in the pretesting session. Subjects were then thanked and dismissed with the information that they would be called when all experimental procedures had been finished "to learn their personal results" and to receive further information about the overall results of the experiment.

Protocol, *a priori* Hypotheses, and Statistical Evaluation

The experimental procedures were agreed on in advance and noted in a protocol deposited with the sponsor before commencement of the experiment. The hypotheses were derived from the literature and stated as specific post-hoc tests to be evaluated in a 2 × 4-way analysis of covariance with the pretest scores as covariates, in case main effects were significant at the .05 conventional level. In particular we hypothesized that

1. There would be a genuine effect of the instruction and accordingly of the expectation to receive caffeine (true placebo effect), operationalized as a significant difference between group b (instruction caffeine) and group d (control).
2. These effects would not be due to any type of physiological or psychological effect of drinking decaffeinated coffee, operationalized as a significant difference between group a (instruction no caffeine) and group d (control).
3. That the expectation of receiving coffee would create a larger effect than the ambivalence condition, operationalized as a significant difference between group b (instruction caffeine) and group c (double-blind instruction).

Additionally, we expected

4. That there would be a main effect between experimenters, and interactions, such
5. that the expectations of the experimenters would mainly show as interactions, operationalized as interactions of the first order differences between groups a and b (instruction no caffeine, instruction caffeine) and group c (double-blind instruction), such that with experimenter 1 (negative expectation) group a would have lower values than the double-blind group, and with experimenter 2 (positive expectation) group b would have higher values than the double-blind group.
6. That there would not be any influence of the negative expectations of experimenter 1 in group b (instruction caffeine), but
7. that the positive expectations of experimenter 2 would show mostly in group b (instruction caffeine).

While hypotheses 1–3 are straightforward extrapolations from studies done so far and could be reasonably expected to hold up under the preconditions that the experimental paradigm would be robust against slight alterations in place, time, and circumstantial changes in what is called a conceptual replication (Schweizer, 1989), hypotheses 4–7 were intended to gather additional new information through this experiment.

Since the design was a completely randomized experimental design and measures would be continuous and quasi-continuous data, we opted for an analysis of covariance as the appropriate statistical method with pretest measures as covariates and post-hoc tests to mirror the apriori hypotheses. The evaluation was done using Statistica Version 5. Data was double checked after entry. All variables were tested for homogeneity of variance and normal distribution, before any tests were run. All variables except the secondary outcome measure of the computer jump-and-run game complied with the preconditions for ANCOVA.

Results

Subjects

171 subjects answered the advertising campaign, of whom 11 did not show up to the scheduled date. One person was not willing to accept the video taping and declined participation. Three more subjects did not fulfill all inclusion or exclusion criteria. This left us with 157 subjects with a mean age of 28.1 years (range 17–57, *SD* 7.6), 66% of which were female ($n = 104$). 6% were academics, 10% civil servants or white-collar workers, 73% were students, and 11% had other professions. 66% ($n = 77$) of the students or slightly less than half of the overall sample were psychology students. 61% of the subjects had no experience with computer games, but 66% regularly worked with a computer, while 28% did not work with a computer at all. 81% of the sample drank 1–4 cups of coffee daily, the others more than 4 cups; 83% of the sample had been drinking coffee at least since 5 years. Subjects were motivated to participate in this experiment mainly by curiosity (32%), by reimbursement (28%), scientific interest (17%); 21% gave other reasons, and only 1% said they were participating out of boredom.

Difficulties

Although the instruments used for measuring blood pressure and heart rate were newly purchased, one of them broke down after a few days and had to be re-

Table 1

Main and secondary variables; mean values (SD) per group and for the total sample expressed in post-pre differences (positive values reflecting an increase in the variable); main variables: heart rate (beats per minute), blood pressure systolic and diastolic (mm/Hg), sum score wellbeing (Basler Befindlichkeitsskala BBS; Hobi, 1986) high scores: better well-being, low scores: worse well-being); secondary variable: number of misprints found in text within 4 minutes; secondary variable "computer jump-and-run game" was not normally distributed and is therefore not presented here, it was evaluated by nonparametric statistics; see text.

	Group A No Caffeine	Group B Caffeine	Group C Double-Blind	Group D Control	Total
N =	41	39	39	38/37	157/156
Heart rate	−6.2 (8.3)	−4.9 (5.9)	−4.5 (6.8)	−4.1 (5.2)	−5.0 (6.7)
Blood pressure (systolic)	−7.9 (7.7)	−6.4 (7.4)	−4.3 (9.2)	−3.5 (2.7)	−5.6 (8.1)
Blood pressure (diastolic)	−2.3 (5.6)	−3.7 (4.3)	−2.3 (5.8)	−2.7 (4.9)	−2.7 (5.2)
Well-being	3.6 (8.1)	−1.7 (11.6)	0.1 (8.8)	−0.8 (9.3)	0.4 (9.6)
Misprint Correction	0.6 (5.2)	0.3 (5.5)	1.0 (4.7)	0.7 (3.5)	0.7 (4.8)

placed. This resulted in data loss for 36 subjects. The data before the breakdown were screened for outliers and the blood pressure analysis was rerun also with windsorized data and data corrected for outliers. After one week the experimenter who was instructed that placebo effects were probably only artifacts started to disbelieve his experimental information and to suppose that placebo probably could be effective. To check whether this had any influence, the data were subsequently recoded according to the newly expressed opinion of this experimenter and the analysis rerun. But results remained essentially the same. Therefore, the results are reported according to the original experimental assignment of experimenters.

Experimental Effects

The main results of the experiment for the instruction factor can be seen in Table 1, for the experimenter factor in Table 2, which both give pre- and postscores together with the standard deviations (*SD*).

Heart rate showed no significant main effects in either factor. Systolic blood pressure showed no significant main effect in the instruction factor, but a highly significant difference in the experimenter factor ($F\,1/147 = 7.01$; $p = .008$), with the group instructed by experimenter "proplacebo" having a significantly higher blood pressure (117.1) than the group instructed by experimenter "antiplacebo" (111.3). There was no significant main effect in the instruction factor on diastolic blood pressure and only a marginally significant effect of the experimenter factor ($F\,1/147 = 3.14$; $p < .08$). Well-being was significantly changed by the instructions ($F\,3/145 = 3.17$; $p < .03$). Post-hoc analysis revealed a highly significant difference ($p < .0004$) between the group "told without caffeine," which was feeling better (79.9), and the control group feeling worse (71.5). There was no significant effect of experimenters on well-being. While there were no significant interactions in these above-mentioned variables, there was a significant interaction in the secondary variable "misprint correction" ($F\,3/145 = 3.09$; $p < .03$), while no main effects in this variable were significant. Subjects instructed by experimenter "proplacebo" did significantly worse than all the other subjects when in the group "caffeine" ($p < .0001$). Since the secondary variable "computer game" was not normally distributed, analysis was done by the nonparametrical Kruskal-Wallis test on instructions and experimenters separately: There were no significant differences.

In order to find out whether expectation had an influence on the result, we correlated the expectations of subjects with the pre/post differences in the outcome

Table 2
Main and secondary variables; mean values (*SD*) per experimenter and for the total sample sample expressed in post-pre differences (positive values reflecting an increase in the variable); main variables: heart rate (beats per minute), blood pressure systolic and diastolic (mm/Hg), sum score well-being (Basler Befindlichkeitsskala BBS; Hobi, 1986) high scores: better well-being, low scores: worse well-being); secondary variable: number of misprints found in text within 4 minutes; secondary variable "computer jump-and-run game" was not normally distributed and is therefore not presented here, it was evaluated by nonparametric statistics; see text.

N = Variable	Experimenter pro-placebo 78/77 Pre	Experimenter anti-placebo 79/78 Pre	Total 157/156 Pre
Heart rate	–5.1 (6.8)	–4.8 (6.6)	–5.0 (6.7)
Blood pressure (systolic)	–4.6 (7.8)	–6.5 (8.2)	–5.6 (8.1)
Blood pressure (diastolic)	–2.5 (6.2)	–3.0 (4.1)	–2.7 (5.2)
Well-being	1.3 (10.0)	–0.4 (9.2)	0.4 (9.6)
Misprint Correction	0.9 (4.8)	0.4 (4.7)	0.7 (4.8)

variables. We found only one notable correlation: The expectation that coffee would affect concentration negatively was significantly correlated with a larger difference pre/post in well-being (Spearman's $R = .23$; $p = .006$), i. e., subjects expecting a lack of concentration tended to feel slightly worse after the intake of the caffeine placebo. Expectation about the effect of caffeine on well-being was not correlated with change in well-being. Correlations between the respective expectations and results broken down according to groups were not sizeable and if significant only slightly so. Considering the multiplicity of tests we are inclined to attribute these small correlations to chance fluctuations.

After all the data were registered, the subjects assessed the subjective probability of having received coffee using a visual analog scale (0 = surely coffee, 10 = surely placebo). We correlated this probability with the difference scores of the outcome variables for all experimental groups except the control group, which had not been asked this question. There were no significant correlations between subjective probability of having received caffeine and any of these outcome variables, the correlations being insubstantial and ranging between .06 and .14. Subjects rated their subjective probability of having received coffee generally according to the experimental manipulation. The respective scores were: group "told placebo" 6.9, group "told coffee" 2.9, group "double blind" 6.0. These scores show that the experimental manipulation had generally been successful, albeit not to the same degree in all subjects. However, subjects in the group "told coffee" instructed by experimenter "proplacebo" rated their probability as 1.4 (*SD* 1.8), while subjects in this group instructed by experimenter "antiplacebo" rated it as 4.2 (*SD* 2.3), which indicates nearly equal probability of coffee and placebo and shows that these subjects were unclear as to whether they had actually received coffee. We calculated a 3 × 2-way ANOVA on subjective probability of having received caffeine (without the control group, which had not been asked this question) and found a highly significant main effect for the instruction factor ($F\ 2/112 = 22.61$; $p < .00001$), a significant main effect for the experimenter factor ($F\ 1/112 = 4.14$; $p < .05$), and a significant interaction ($F\ 2/112 = 3.27$; $p < .05$). This analysis confirms that the experimental manipulation had been successful, that there were differences between the two experimenters, and that the experimenters affected subjects selectively. While the difference was not very large, in the group instructed "no caffeine" and in the double-blind group the caffeine instruction was not successful with experimenter antiplacebo. Figure 1 depicts this situation graphically.

Debriefing Interviews: Impression of Experimenters and Plausibility of Experiment

After all data had been deposited, the subjects were contacted by phone and informed "about the results of the experiment" as promised. They were informed about the true object of the study, about the fact that no caffeine was employed, and were asked about the plausibility of the experiment as a whole and how they rated some aspects of the experimenters, like getting along with them, clarity in introducing the experimental task, self assured attitude, attractiveness, each item being rated on a three-point rating scale.

On the whole, subjects believed the cover story of the experiment: 50% found it plausible, 26% somewhat

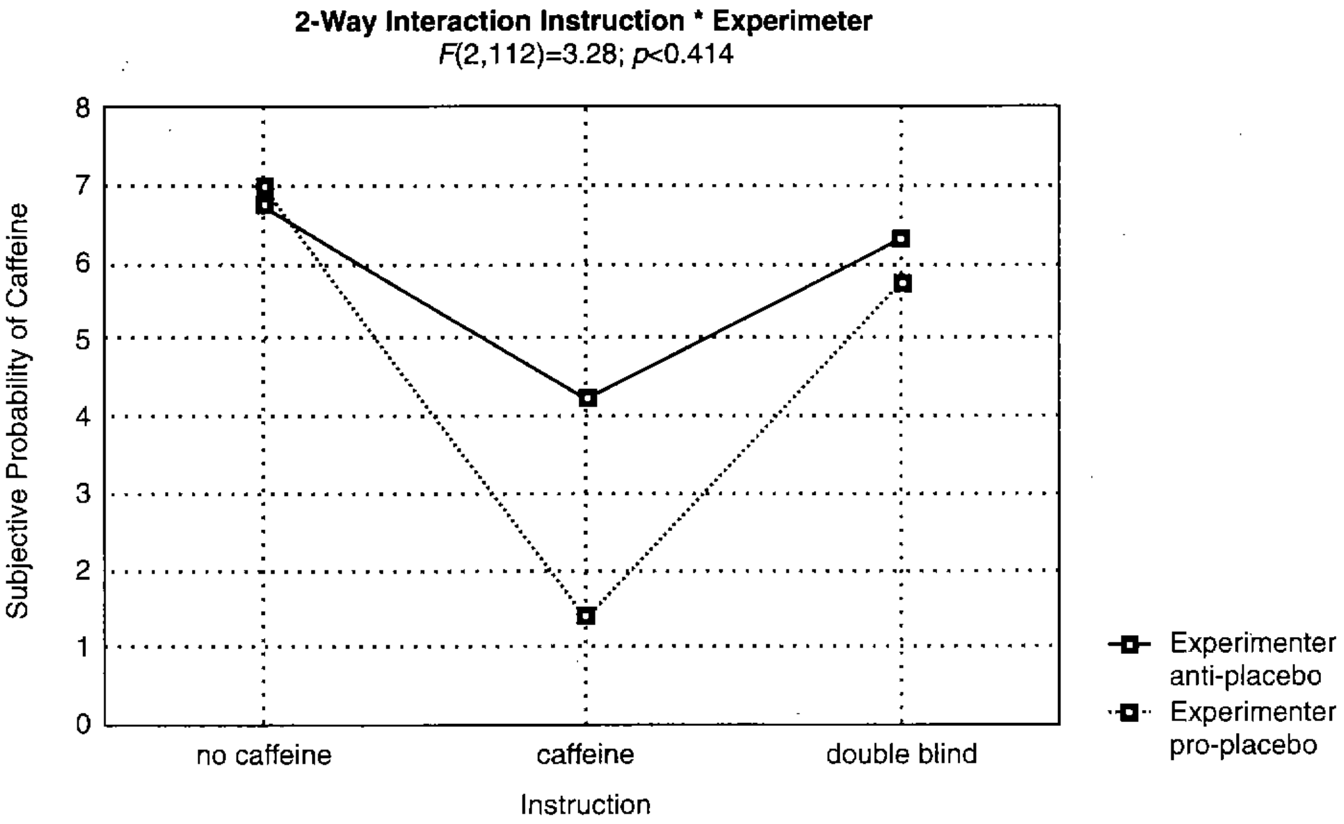

Figure 1
Subjective probability of having received caffeine (VAS, 0 = certainly caffeine; 10.5 = certainly no caffeine) according to experimental groups.

plausible, and 15% thought it implausible, with 8% missing data. 76% of the subjects said they had no idea that only decaffeinated coffee had been served, 15% mentioned they had discovered the fraud, 8% of the data were missing. Interestingly, subjects instructed by experimenter "antiplacebo" were significantly more suspect about the caffeine used (Mann-Whitney $p = .001$, two-tailed). Experimenter "proplacebo" was rated significantly better (Mann-Whitney $p < .001$, two-tailed) in the variables "getting along," "self-assured," "attractiveness." No difference was found in the way they introduced the experimental task or in the plausibility of the experiment.

Discussion

This experiment used a standard experimental procedure to induce expectancy effects. Our aim was to study how these expectancy effects are mediated and to find a clue by manipulating experimenters' expectancy. Experimenters were hired and totally blind as to the true experimental question and setup. When debriefed, experimenters were astonished about the fact that only decaffeinated coffee had been used. The junior authors who knew about the design and purpose of the experiment had no contact with experimental subjects, nor had the senior authors had any contact. The debriefing of subjects and experimenters took place only after all original data had been deposited with the senior authors. No comparable experiments have ever been done in Freiburg. The decaffeinated coffee was not detected as such in a blinded informal pretest. We therefore are confident that the blindness of the experiment was maintained throughout the experiment. The debriefing interviews of the subjects confirm this. Although the data of 13 persons (8%) were not available, three-quarters of the subjects had no idea that only decaffeinated coffee was being used and believed the cover story. Since student subjects are sometimes aware of the fact that psychological experiments have covert purposes and may therefore suspect some kind of fraud anyway, we take this rate to be good evidence for the fact that subjects generally believed in what they were told. Furthermore, subjects were in no way dependent on either experimenters or any of the authors for credit or other benefits. They mostly participated for curiosity's sake or for the money offered as a reward. We therefore feel that the results are externally and internally valid.

Thus, it is interesting to note that we were not able to replicate the findings of previous studies using similar approaches showing clear effects of expectancy. None of our primary or secondary variables showed any effect of the instructions in the hypothesized way. There was one effect in the opposite direction: Subjects told that they would receive a placebo felt significantly better than subjects of the control group. Since the main effect was significant at $p < .03$, and we ran six independent multiple analyses: This significance is in fact lost within the multiplicity of tests. We therefore are inclined to dismiss this theoretically contradictory effect as a chance finding. The only effect faintly pointing into the predicted direction is the significant interaction between the instruction group "told caffeine" when instructed by experimenter "proplacebo" and all the other groups in the misprint correction task. These subjects did significantly worse than all other subjects. However, this runs contrary to what one would expect from the coffee stereotype, namely a better performance in concentration tasks. Rather the performance in the coffee instruction group instructed by experimenter "proplacebo" is significantly worse. Apart from this effect, we could, in sum, not reproduce the findings of other studies (Fillmore & Vogel-Sprott, 1992; Fillmore et al., 1994; Kirsch & Weixel, 1988) that instructions and expectancies alter blood pressure, heart rate, general well-being, or cognitive performance.

This was not because of a lack of power, since our study used twice as many subjects as the studies reported in the literature. Two explanations stand out: None of the studies were ever done in Germany, apart from an early one (Lienert, 1955), which explicitly mentioned "suggestion." It could be the case that the caffeine stereotype is not as strong in Germany as in the United States and Canada. Furthermore, subjects might not have had a uniform expectation and were not supplied one by us, since we trusted the general coffee stereotype to work also in a new setting. This obviously is not the case, as is also shown by the lack of significant correlations of expectations with effects. Without a clear rule as to what effect can possibly be expected, provided either by the social stereotype or by some kind of information, instructions alone apparently are not strong enough to induce an expectancy.

One might argue that people just did not believe in any kind of experimental manipulation at all, so that no expectancy whatsoever could result. Suspicion has been voiced as to the validity of balanced placebo designs in general, since they might be more reflective of the desire to please the experimenter than of true expectancies (Knight et al., 1986). While this argument might be true for some of the studies reported in the literature, it does not seem to be a plausible explanation for our results: We double checked the success of our manipulation, by asking about the probability of having received coffee immediately after the experiment and about 3 weeks later in a debriefing interview. ANOVA revealed a highly significant instruction effect, and the descriptive statistics show that generally subjects believed the experimental information, as was confirmed in the debriefing interview. Subjects at that time had no reason to please anybody, because they were told the truth and then asked whether they knew it all along. If judged conservatively, a quarter of the subjects might have suspected some aspect of the true design, but this would not be enough to invalidate the whole data set.

The most plausible explanation for our lack of expectancy effects, apart from a general chance fluctuation, is the weakness and diversity of the social stereotype of caffeine effects. We tested this hypothesis in a second study using the same design by varying the stereotype and found similar results (data to be reported). At least in Germany there does not seem to exist a uniform stereotype of what coffee is supposed to do. Coffee seems to be perceived more as a general purpose drink than as a drug with specific effects.

Although the main precondition for studying the influence of experimenter expectation on subject expectancy was not met, we found two clear-cut effects of experimenters: Subjects instructed by experimenter "proplacebo" had a significantly ($p = .008$) higher systolic blood pressure than subjects instructed by experimenter "antiplacebo." Since this main effect and the according post-hoc test are much stronger than the above-mentioned ones, we take this to be a genuine effect of the experimenter. Experimenters also had different success in conveying the belief that subjects had consumed coffee, with experimenter "antiplacebo" generally being less effective and mainly so in the group instructed to consume coffee. Note that both experimenters were equally blind to the fact that no caffeine was being used. This is an interesting finding, for the experimental instruction seems to work in two ways. Experimenters were generally different in conveying a sense of the effect to be expected—main effect in blood pressure—and selectively on those subjects whom we wanted to expect coffee (interaction). It is therefore safe to conclude that the main trigger of subjects' expectancies are experimenter expectations, which modulate the experimental instructions accordingly. It is important to bear in mind that experimenters actually had changed their opinion as to the placebo effect. Experimenter "proplacebo" continued to believe the experimental information that "the placebo effect is a real effect," albeit it in a weaker version toward the end of the experiment. Experimenter "antiplacebo" had changed his mind after 7 days or one-third of the experiment. He discontinued to believe that "placebo effects are artifacts" and instead thought that placebos could have real effects after all. As was mentioned, another analysis taking into account these "real" expectations of experimenter "antiplacebo" did not change the picture. Results basically remained the same. This could mean that the experimental information was still in a way active although the experimenter reported a different opinion.

Apart from that, experimenters differed in a number of ways, which were assessed by a personality profile and by asking subjects at the debriefing interviews about the experimenters. Experimenter "proplacebo" was extremely satisfied with his life, relatively restricted in social relations, had very few bodily complaints and concerns about his health. He described himself as rather introverted and his neuroticism score was low. Experimenter "antiplacebo," on the other hand, was average in nearly all of the scales, or his scores were comparable with those of experimenter "proplacebo," except for the scale "social orientation," which was very high and oriented toward a more altruistic stance. On the video tapes, experimenter "antiplacebo" exhibits a rather dispassionate and detached way of relating, while experimenter "proplacebo" is more active and enthusiastic about his job, a clear discrepancy with their self-descriptions in the personality scales. Accordingly, subjects rated experimenter "proplacebo" generally more favorable: They found him significantly more attractive, more self-assured, they got along with him better, and with him they said less that they had a hunch that no caffeine was being employed. Clearly, experimenter "proplacebo" was the favorite. He obviously intrigued subjects more than his colleague did, and this might be explanation enough for the higher arousal in these subjects, for less suspicion, and for the greater conformity he aroused about the experimental task. Note that there was no difference in the way the experimenters explained the procedures. Both were rated as clear. It seems that interpersonal perception and affect is more important than the more cognitively directed explanations and expectations of the experimenter. Therefore, it seems plausible to conclude that there were differences in the way experimenters affected subjects' arousal and expectations of what they would receive as a beverage. These differences were probably due to personal differences in the experimenters and not to experimental manipulations or information and in consequence they were not due to specific expectations experimenters might have had.

One possible way for future studies to study experimenters' expectations could be to use persons who naturally and strongly adhere to one or another set of beliefs, document them, and see whether they succeed in eliciting differential effects in subjects. The problem could thus be circumvented as to how to stably install beliefs in experimenters.

This study is to our knowledge the first within this research paradigm to employ experimenters who were ignorant about the hypothesis being studied and completely blind as to the research questions. It could well be the case that expectancy effects reported in the literature are in fact experimenter effects from the fact that experimenters were *not* completely blind to research questions, hypotheses, and expected results, but only to actual group assignments. We would therefore like to see replications of expectancy effects using more rigorous blinding procedures in order to rule out any type of experimenter effects.

Our objective to demonstrate an instruction effect and to study the way this is mediated failed, except for one interaction effect: Subjects instructed by experimenter "proplacebo" did worse in the misprint correction task. They probably relied on the coffee, which was not supplied, and concentrated less. We could not demonstrate other effects, contrary to other findings in the literature. Our inability to replicate results reported in the literature might be due to the fact that the stereotype of what coffee is supposed to do might not be so clear-cut in Germany as it is in the United States and Canada. We hope that this will prove to be a plausible explanation for our results and that lack of blinding in other studies is not the case at issue, after all.

Acknowledgments

This work was supported by a grant from the Institut für Grenzgebiete der Psychologie, Freiburg. We are grateful to Prof. Fahrenberg, Freiburg, for helpful advice. This work was carried out as a diploma thesis in psychology by YB and SW.

References

Camí, J., Guerra, D., Ungena, B., Segura, J., & de la Torre, R. (1991). Effect of subject expectancy on the THC intoxication and disposition from smoked hashish cigarettes. *Pharmacology and Biochemistry of Behavior, 40*, 115–119.

Fahrenberg, J., Hampel, R., & Selg, H. (1989). *Das Freiburger Persönlichkeitsinventar FPI. Revidierte Fassung FPI-R und teilweise geänderte Fassung FPI-A1* [Freiburg Personality Inventory FPI. Revised Version FPI-R and Partially Changed Version FPI-A] (5th ed.). Göttingen: Hogrefe.

Fillmore, M., & Vogel-Sprott, M. (1992). Expected effect of caffeine on motor performance predicts the type of response to placebo. *Psychopharmacology, 106*, 209–214.

Fillmore, M.T., Mulvihill, L.E., & Vogel-Sprott, M. (1994). The expected drug and its expected effect interact to determine placebo responses to alcohol and caffeine. *Psychopharmacology, 115*, 383–388.

Hobi, V.(1985). *Basler Befindlichkeits-Skala. Ein Self-Rating zur Verlaufsmessung der Befindlichkeit* [Basle Well-Being Scale. A selfrating scale for time course measurement of well-being]. Weinheim: Beltz Test Gesellschaft.

Hughes, J.F., Gulliver, S.B., Amori, G., Mireault, G.C., & Fenwick, J.F. (1989). Effects of instructions and nicotine on smoking cessation, withdrawal symptoms and self-administration of nicotine gum. *Psychopharmacology, 99*, 486–491.

Hull, J.G., & Bond, C.F. (1986). Social and behavioral consequences of alcohol consumption and expectancy: A meta-analysis. *Psychological Bulletin, 99*, 347–360.

Kienle, G.S. (1995). *Der sogenannte Placeboeffekt. Illusion, Fakten, Realität* [The so-called placebo effect. Illusions, facts, reality]. Stuttgart: Schattauer.

Kienle, G.S. (1996). Der Placeboeffekt—Realität oder Illusion [The placebo effect—Reality or illusion]. In J. Hornung (Ed.), *Forschungsmethoden in der Komplementärmedizin. Über die Notwendigkeit einer methodologischen Erneuerung* [Research methods in complementary medicine. On the necessity of a methodological renewal] (pp. 84–97). Stuttgart: Schattauer.

Kienle, G.S., & Kiene, H. (1996). Placebo effect and placebo concept: A critical methodological and conceptual analysis of reports on the magnitude of the placebo effect. *Alternative Therapies in Health and Medicine, 2*, 39–54.

Kirsch, I. (1994). Clinical hypnosis as a nondeceptive placebo: empirically derived techniques. *American Journal of Clinical Hypnosis, 37*, 95–106.

Kirsch, I. (1997). Specifying nonspecifics: Psychological mechanisms of placebo effects. In A. Harrington, *The placebo effect: Interdisciplinary explorations* (pp. 166–186). Cambridge, MA: Harvard University Press.

Kirsch, I., & Rosadino, M.J. (1993). Do double-blind studies with informed consent yield externally valid results? An empirical test. *Psychopharmacology, 110*, 437–442.

Kirsch, I., & Weixel, L.J. (1988). Double-blind versus deceptive administration of a placebo. *Behavioral Neuroscience, 102*, 319–323.

Knight, L.J., Barbee, H.E., & Boland, F.J. (1986). Alcohol and the balanced placebo design: the role of experimenter demands in expectancy. *Journal of Abnormal Psychology, 95*, 335–340.

Lienert, G.A. (1955). Die Bedeutung der Suggestion in pharmakopsychologischen Unterschungen [The importance of suggestion in pharmacopsychological studies]. *Zeitschrift für experimentelle und angewandte Psychologie, 3*, 418–438.

Marlatt, G.A., & Rohsenow, D.J. (1980). Cognitive processes in alcohol use: Expectancy and the balanced placebo design. In N. Mello (Ed.), *Advances in substance abuse (1)* (pp. 159–199). Greenwich: JAI Press.

Schweizer, K. (1989). Eine Analyze der Konzepte, Bedingungen und Zielsetzungen von Replikationen [An analysis of concepts, prerequisites, and goals of replications]. *Archiv für Psychologie, 141*, 85–97.

Zwyghuizen-Doorenbos, A., Roehrs, T.A., Lipschutz, L., Timms, V., & Roth, T. (1990). Effects of caffeine on alertness. *Psychopharmacology, 100*, 36–39.