



ELSEVIER

International Journal of Psychophysiology 43 (2002) 247–260

INTERNATIONAL
JOURNAL OF
PSYCHOPHYSIOLOGY

www.elsevier.com/locate/ijpsycho

The effects of a caffeine placebo and suggestion on blood pressure, heart rate, well-being and cognitive performance

Harald Walach*, Stefan Schmidt, Thomas Dirhold, Sonja Nosch

Universitätsklinikum Freiburg, Institut für Umweltmedizin und Krankenhaushygiene, Hugstetterstrasse 55, 79106 Freiburg, Germany

Received 15 February 2001; received in revised form 21 July 2001; accepted 26 July 2001

Abstract

We studied the effect of suggestion and different instructions in a balanced placebo design. One hundred and fifty-nine subjects were randomized into a 2*4 factorial design. All subjects except a control group received a caffeine placebo. Subjects were randomized to a condition which was identical to an earlier study, or received an information about scientifically proven effects of caffeine (factor 1). The second factor varied instructions: subjects were either made to expect coffee, no coffee or were in a double blind condition and were told either coffee or placebo would be applied. Dependent measures were blood pressure, heart rate, well-being and a cognitive task. There was one main effect of the instruction factor ($P = 0.03$) on diastolic blood pressure, with the group 'told caffeine' reporting significantly smaller decrease in diastolic blood pressure than controls and subjects in the double blind condition. There were no other main effects on both the instruction or suggestion factor, and no interactions. Contrary to the literature, instruction effects were very small. This was apparently due to the fact that placebo-caffeine in the dose used in this study — one cup of strong 'coffee' — did not produce expectancy effects strong enough to affect the parameters measured. It is concluded that the placebo-caffeine research paradigm is not suitable for researching instruction effects in Germany, and that reported effects should be reproduced with tighter controls. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Caffeine; Balanced placebo design; Instruction; Expectation; Placebo effect

* Corresponding author. Tel. +49-761-270-5497; fax +49-761-270-7224.

E-mail addresses: walach@ukl.uni-freiburg.de (H. Walach), sschmidt@ukl.uni-freiburg.de (S. Schmidt).

1. Introduction

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 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 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. Subjects expecting to consume alcohol but actually receiving an alcohol-placebo tend to be more aggressive, more sexually aroused, consume more drinks and experience more craving, while the actual intake of the equivalent of 0.1% alcohol does not produce these effects. Cognitive and behavioral measures, however, do not seem to be affected by expectations, but by the intake of the substance itself. This result has been substantiated by a meta-analysis (Hull and 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 $g = 0.08$; $z = 2.03$; whereby g is Hedge's g , a difference of means standardized with the pooled standard deviation and corrected for study size (Hedges and Olkin, 1985). Interestingly, no interactions occurred.

Studies on the effects of expectancies have been also carried out in other areas (Camí et al., 1991; Hughes et al., 1989; Kirsch and Rosadino, 1993; Fillmore and Vogel-Sprott, 1992; Fillmore et al., 1994; Kirsch and Weixel, 1988; Lienert, 1955; Murray, 1988; Zwyghuizen-Doorenbos et al., 1990). In all of these studies substantial effects of expectancies have been demonstrated. The studies of Fillmore and Vogel-Sprott (1992) and Fillmore et al. (1994) followed up the suggestion of Hull and Bond (1986) and used only one-half of the balanced placebo-design to study the effect of expectancy of receiving caffeine: apart from a natural history control group with no

experimental manipulation all groups received placebo-coffee. The first of these studies showed a significant effect of caffeine expectation on motor function and well-being in subjects who were instructed to expect caffeine. The second experiment studied the interaction of drug-expectations and contradictory instructions. Two groups of subjects received a caffeine-placebo and the instruction, coffee would either 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 and 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 than instructions geared towards the opposite direction.

Kirsch and Weixel (1988) used the same approach, with a double-blind group added, 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. The measures were a rotor pursuit task, reaction time, finger span, symbol substitution, heart rate and blood pressure, well-being, as well as the subjective probability of having received 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 of having received caffeine. Moreover, there were dose-dependent effects. While in the open instruction group the effects were initially stronger with higher doses and then leveled out, in the double-blind condition effects were stronger with higher doses. This led the authors to question the validity of the double-blind paradigm for

evaluating drugs in clinical trials. Lienert, using a balanced placebo design already in 1955 in what was probably the earliest application of this design, found that positive suggestions of having received caffeine could alter heart rate (Lienert, 1955).

Taken together, these studies seem to suggest:

1. a fairly clear-cut and consistent effect of the expectancy to receive caffeine, even if no caffeine is employed; caffeine expectancy seems to improve well being, alertness and cognitive functioning, raise blood pressure and heart rate;
2. differential effects of double-blind administration of a caffeine-placebo; and
3. significant effects against zero-controls in both psychological and physiological variables, and in behavioral variables as well.

Starting from these results, we had, in an earlier study (Walach et al., 2001), tried to reproduce these results in a 2*4-factorial design with 157 subjects. All subjects, except a control group, received placebo-coffee. Two experimenters were led to expect placebos either to produce physiological effects or not (pro- vs. anti-placebo). Subjects were either made to expect coffee, no coffee or were in a double blind condition and were told either coffee or placebo would be applied. Dependent measures were blood pressure, heart rate, well-being and a cognitive task. There was one main effect of the instruction factor ($P = 0.03$) with the group 'told no caffeine' reporting significantly better well-being. There was one main effect of the experimenter factor with subjects instructed by experimenter 'pro-placebo' having higher systolic blood pressure ($P = 0.008$). There was one interaction with subjects instructed by experimenter 'pro-placebo' to receive coffee doing worse in the cognitive task than the rest. Subjects instructed by experimenter 'anti-placebo' were significantly less likely to believe the experimental instruction and mostly if they had been instructed to receive coffee. Contrary to the literature we could not show an effect of instruction, but an effect of experimenters. This effect, however, was due to personality differences, and not to the

information given to experimenters. In order to reproduce our own findings which contradict much of the literature, and also to analyze this failure to reproduce in more detail, we conducted another study with basically the same procedures and design. In order to keep conditions as similar as possible, we also used two hired experimenters who were blind to the real purpose of the study and to research hypotheses, who collected the data and had contact with subjects. We used the same instruction conditions as in the previous study. Instead of varying information of the experimenters both experimenters received the same information. However, we introduced another factor: replication vs. suggestion. In one condition — replication — subjects did not receive additional information. In the suggestion condition, subjects received an additional information leaflet suggesting scientifically proven effects of caffeine.

According to the expectancy model of Kirsch (1985, 1997), placebo effects are induced by expectancies, which develop in subjects due to their learning history and to explicit expectancy manipulations. Our aim was two-fold with this study: We wanted to see whether the null results of the previous study (Walach et al., 2001) were themselves reproducible. For that reason we adhered as closely as possible to the previous protocol. Also we wanted to find out whether the lack of expectancy effects in the caffeine placebo model is possibly due to a lack of potent expectancies concerning caffeine in the first place. Therefore, we measured specific expectancies in this study.

2. Method

2.1. Subjects

In order to avoid leakage of information due to the first study, we conducted the whole experiment on the premises of the local police-academy which has nothing to do with the university. Subjects were mainly recruited from this academy, and among medical students, who had no contact with subjects of the previous study. Subjects were approached by flyers, posters, in classrooms and

by word of mouth. They were either informed in groups or were informed by telephone about the study. On that occasion they also learned about the inclusion and exclusion criteria for participating in the study. These were:

inclusion criteria:

- regular consumption of coffee (at least one cup per day)
- minimum age 16

exclusion criteria:

- psychiatric diagnosis
- circulation and 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 h before the experiment.

One hundred and eighty-six subjects answered the advertising campaign. Twelve were used for pre-trial runs, and 14 did not show up at the scheduled date. One person was excluded because she was not a habitual coffee drinker (at least one cup per day). This left us with 159 subjects with a mean age of 25.5 years (range 16–60, S.D. 7.6), 42% of which were female ($n = 67$). Randomization was successful in creating groups with comparable gender distribution. The proportion of females varied between 38% and 50% in the experimental groups (see Section 2.2). Sixty-two percent were students at the police academy, 2.5% were policemen in training, 9% were university students, 4% were physiotherapists and nurses, 21.5% were members of diverse groups (high-school students, self-employed, workers, white collar workers, out of work).

Thirty-two percent of the sample drank up to one cup of coffee a day, 56% drank between one and four cups a day, 11% drank more than 4 cups, with 2 persons having missing data. Seventy-five percent of the sample had been drinking coffee at least since 2 years. Subjects were motivated to participate in this experiment mainly by curiosity (47%), by reimbursement

(19%), scientific interest (20%), 16% gave other reasons and only 1% said they were participating out of boredom.

2.2. Design

We used a 2*4-factorial design. The first factor represented the conditions 'Replication' or 'Suggestion', the second factor the instructions and experimental manipulations of the subjects.

2.2.1. Replication vs. suggestion

The first factor represented the way how the replication was carried out. It had two levels: Level one was a pure replication of the first study without any further intervention. Subjects in level one underwent the experiment without further information. Level two introduced an added suggestion of what to expect from caffeine. This we tried to achieve by giving subjects an information leaflet, which described effects of caffeine according to pharmacological knowledge, e.g. increasing blood pressure, activation, etc. This was meant to increase expectation of caffeine effects.

2.2.2. Subject expectancy

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 (Deutsch Mark 25 for 1.5 h), and after having given written informed consent they were randomly allocated to one of four groups:

1. true information: subjects received a caffeine-placebo and were told so;
2. false information: subjects received a caffeine-placebo, but were told that they would receive coffee;
3. 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;
4. natural history control: no substance and no instruction was given.

2.3. Experimenters

Two experimenters were recruited from a number of applicants who had answered posters on the campus. They were both male, of the same age, and comparable in their appearance. They were reimbursed (14 Deutsch Mark per hour) and instructed that the purpose of the study would be to study the placebo-effect in a setting, where both coffee and a placebo was used to purportedly assess 'effects of caffeine on working at the computer screen'. We made sure in an introductory interview that they did not know anything about placebo-effects and expectancy research. While in the predecessor study we varied the information given to the experimenters, in this study we employed them to achieve complete blinding of the experiment, apart from replicating the experimental set-up of the previous experiment. Experimenters were trained thoroughly by the junior authors (T.D. and S.N.) in all experimental procedures, including the use of the measuring devices. They were trained separately and did not meet with each other, neither did they incidentally know each other. Each experimenter did some trial runs with pilot subjects to make sure that they had understood the procedure. Experimenters were completely blind as to the real purpose of the experiment and as to the fact that only decaffeinated coffee was being used. They were located in separate buildings, and did not meet each other. During half of the time, they switched rooms. 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. They did not interfere with the experimental subjects and had a separate room where they could observe the experimental procedures on a video screen. All experimental sessions were videotaped.

2.4. Measures

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

1. Blood-pressure (systolic and diastolic) using an automatic measuring device ('boso-carat', by Bosch & Sohn, Jungingen, Germany) which automatically inflates and deflates the cuff and displays the reading on a display. This is a device frequently used in general medical care routine and self-monitoring of blood pressure. Measurements are accurate with a mean error of 0.98 mmHg systolic and 3.2 mmHg diastolic, according to an independent expertise of the Technical University of Dresden by appointment of the German Federal Institute for Physical and Technical Standards (unpublished material on request from the manufacturer). Two readings were taken 3 min apart from each other and averaged to yield blood-pressure scores.
2. Heart rate using the same device, simultaneously with the blood pressure.
3. General well-being using the Basler Befindlichkeitsskala (Hobi, 1985), an adjective list which measures general well-being in 16 pairs of adjectives with opposite semantic content arranged as a semantic differential. It measures the four dimensions alertness, extraversion, vitality, inner balance and can also be summed to one general dimension. We used this scale, because it is virtually identical to that used by Kirsch and Weixel (1988), thus allowing comparison. Apart from that it is one of the frequently used validated and standardized German language scales for the measurement of short term changes in well-being, specifically with repeated measures (Westhoff, 1993). To enhance credibility of the experiment, we measured as secondary dependent variables, which to the subjects, however, appeared to be the primary outcome variables, both before and after experimental manipulation.
4. Performance in a self-devised test; the test consisted in finding misprints in a text which was displayed as a word-document on the screen for 4 min. For pre- and post-test 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.
5. Performance in a self-devised computer test 'Click'. In this test, the computer screen displayed a matrix of 13*13 buttons, each of which showed a combination of three letters (s, y, z, x) in random order. Some of the combinations displayed an 'x'. These had to be clicked with the mouse to collect scores. Clicking the wrong buttons resulted in score reduction. Subjects had 2 min per matrix to find the wrong buttons, and two screens were displayed per test. This latter test was the only alteration compared to the predecessor study, in which we had used a fun computer game. Since this game had a very skewed distribution and showed a drastic training effect, we devised this test, which also was used to make the experiment appear more like a real computer-test experiment. The new test showed only a slight training effect and the scores were normally distributed. Apart from these dependent variables we also assessed some moderator variables before or after the experiment.
 6. Expectations: (i) specific expectations about the effects of drinking one cup of coffee (only experimental groups a–c; control group subjects indicated what they thought would have happened, had they drunk a cup of coffee) on all variables a–e, after the pre-test and group allocation, and after consumption of the beverage on a 5-point, balanced Likert scale; (ii) subjective probability of caffeine intake after the post-test on a visual analogue scale with the anchors 'surely no caffeine' and 'surely caffeine'; this also served the purpose of manipulation control.
 7. Descriptive data: demographic data, motivation in participation, coffee drinking behavior, experience with computer.
 8. Inclusion- and exclusion criteria (see above) with a questionnaire at the beginning of the experiment and as part of the procedure of seeking written informed consent.

After the whole experiment was over and all data deposited, subjects were debriefed in a tele-

phone 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:

- expectations;
- quality of interaction with the experimenter;
- their perception of the experimenter (agreeability, ability to convey the experimental tasks);
- being aware of the video camera;
- arousal before and during experiment;
- the plausibility of the procedures and the overt experimental question (effect of caffeine on computer screen performance); and most importantly
- subjects' own hypotheses about the experiment and whether they had seen through the purpose of the experiment and the deception.

2.5. Procedures

Subjects were informed about inclusion and exclusion criteria and given a subject information leaflet. Interested subjects were scheduled for participation. We avoided recruiting psychology students. Subjects were completely blind to purpose, substances tested and the design of the experiments.

We prepared 80 opaque envelopes for each experimenter according to a random number code using the random-procedure of SPSS, which contained the experimental instructions to the subjects, balanced for the four experimental conditions, and randomly containing the instruction to hand out the additional information leaflet to subjects in the 'suggestion' condition. The envelopes were numbered and opened sequentially.

Subjects were scheduled by the junior authors, such that they would not meet each other. Arriving subjects were greeted by the experimenter. They were again instructed about the experiment, 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. Then they filled in the questionnaires. After that,

blood-pressure and heart rate readings were taken, the second blood-pressure reading was initiated after approximately 3 min. 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 given the consumption instructions, the experimenter left the subjects to themselves to consume the beverage.

The coffee was a widely used decaffeinated coffee from Costa Rica. In a pre-trial run we had made sure that the placebo-coffee resembled real coffee in taste, appearance and smell. To test this, our whole research group unknowingly was provided with this brand of coffee for approximately 10 days, in a single blind informal experiment, without anybody noticing the difference or commenting on the altered quality of coffee. The dosage used was always three heaped spoonful of coffee. Coffee was brewed by a filter coffee-brewing machine. Subjects had 5 min to drink the coffee, black, without sugar; artificial sweetener was provided on demand. No coffee whitener or milk was allowed. Subsequently, subjects were required to wait for 5 min more, to 'allow the caffeine to take effect', while reading a travel magazine.

After this waiting period, subjects filled in the expectancy questionnaire, and then the series of post-tests followed in the same order as in the pre-testing session. Subjects were thanked and dismissed with the information that they would be telephoned when all experimental procedures had been finished 'to learn their personal results' and to receive further information about the overall results of the experiment.

2.6. Hypotheses

Our hypotheses were derived from the literature and stated as specific post-hoc tests (contrasts between specific conditions of factor 1) to be evaluated in a 2*4-way analysis of covariance with the pre-test scores as covariates, in case main effects would be significant at the 0.05 con-

ventional level. In particular we formulated the following contrasts and hypothesized that:

1. there would be a genuine effect of the instruction to receive caffeine (true placebo effect), defined as a significant difference between group b (instruction caffeine) and group d (control), i.e. contrast b–d;
2. these effects would not be due to any type of physiological or psychological effect of drinking decaffeinated coffee, e.g. due to taste conditioning, defined as the lack of a significant difference between group a (instruction no caffeine) and group d (control), i.e. contrast a–d;
3. that the expectation to receive coffee would create a larger effect than the ambivalence condition, defined as a significant difference between group b (instruction caffeine) and group c (double-blind instruction), i.e. contrast b–c. Additionally, we expected:
4. a main effect on factor 2 with subjects in the suggestion-condition showing stronger effects than in the no suggestion condition; and
5. positive correlations between subjects' specific expectations of changes in the dependent variable and the measured changes in these variables.

Hypotheses 1–3 are straightforward extrapolations from studies done so far and could be reasonably expected to hold 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 and 5 were meant to analyze purported effects in more detail.

2.7. Protocol and statistical evaluation

The experimental procedures were agreed on in advance, noted in a protocol which was deposited before commencement of the experiment.

Since the design was a randomized experimental design and measures would be continuous and quasi-continuous data we opted for analysis of

covariance as the appropriate statistical method with pre-test measures as covariates and post-hoc tests to mirror the a priori hypotheses. The design was powered to detect a medium-sized effect with a reasonable power. The evaluation was done using SPSS, Version 7. Data were double-checked after entry. All variables were tested for homogeneity of variance and normal distribution, before any tests were run. All variables except expectancies complied with the preconditions for ANCOVA. Correlations between expectancy and actual changes were calculated as Spearman's rho. Differences in expectancies across groups were analyzed by non-parametric Kruskal–Wallis analysis. Control of manipulation was conducted by analyzing belief of caffeine consumption across groups with ANOVA.

3. Results

3.1. Experimental effects

The main results of the experiment for the instruction-factor can be seen in Table 1, for the suggestion-factor in Table 2, which both give

means post scores adjusted for pre-test scores. Table 3 gives the results of the analysis of covariance.

There was no significant main effect on the suggestion factor, and only one significant main effect on the instruction factor: Diastolic blood pressure exhibited a main effect on this factor ($F = 2.98$; $P = 0.03$). Blood pressure was reduced from pre- to post-measurement. The least reduction was seen in the group 'told caffeine'. All a priori contrasts were significant. There were no significant interactions, except a marginally significant interaction ($F = 2.6$, $P = 0.053$) for well-being. Well-being was better in the group 'told caffeine' when a suggestion was added. The interaction plot is presented in Fig. 1.

Expectancy of change was not significantly correlated with actual change, except for expectancy about changes in the computer test 'Click' and results ($\rho = 0.24$; $P = 0.008$). This correlation was totally due to a substantial correlation of these variables only in the double-blind group ($\rho = 0.53$; $P = 0.001$). Results of these correlational analyses between expected change and actual change are presented in Table 4. No interaction effects were seen in these correlations.

Table 1

Main and secondary variables; adjusted post-test means per group and for the total sample pre- and post-experimental manipulation (told coffee, told no coffee, told double blind, control); main variables: heart rate (beats/min), blood pressure systolic and diastolic (mm/Hg), sum score well-being (Balsler Befindlichkeitsskala BBS; (Hobi, 1985); high scores: better well-being, low scores: well-being worse); secondary variables: number of misprints found in text within 3 min; computer test 'Click'

	Group A No caffeine 39	Group B Caffeine 40	Group C Double-blind 40	Group D Control 40	Total 159
Variable	Adjusted post	Adjusted post	Adjusted post	Adjusted post	Adjusted post
Heart rate	72.0	73.4	71.7	72.8	72.5
Blood pressure (systolic)	123.3	123.4	122.1	121.7	122.6
Blood pressure (diastolic) ^a	70.5 ^b	70.7 ^{b,c}	68.0	67.8	69.3
Well-being	80.3	81.9	80.8	80.7	81.0
Misprint correction	30.8	30.2	31.4	31.3	30.9
Computer test 'Click'	38.3	39.4	38.4	38.7	38.7

^aMain effect ($F = 3.32$; $P < 0.05$).

^bSignificant ($P < 0.05$) a priori contrast against the control group.

^cSignificant a priori contrast ($P < 0.05$) against the double-blind group.

Table 2

Main and secondary variables; adjusted post-test means per group and for the total sample pre and post experimental manipulation (factor 2: replication vs. suggestion); main variables: heart rate (beats per minute), blood pressure systolic and diastolic (mm/Hg), sum score wellbeing (Balsler Befindlichkeitsskala BBS; Hobi, 1985; high scores: better well-being, low scores: well-being worse); secondary variable: number of misprints found in text within 3 min; computer test 'Click'

	Replication	Suggestion	Total
<i>N</i>	79	80	159
Variable	Adjusted post	Adjusted post	Adjusted post
Heart rate	72.7	72.3	72.5
Blood pressure (systolic)	122.3	123.0	122.6
Blood Pressure (diastolic)	68.9	69.6	69.3
Well-being	80.0	81.9	81.0
Misprint correction	30.9	31.0	30.9
Computer test 'Click'	38.7	38.7	38.7

No significant main effect or interactions.

Therefore we present only part of the correlations in order to give an impression.

We analyzed the expectancies for change in specific variables according to experimental group. In general, response expectancies were close to indifference (3 = no change expected), ranging from a mean of 3–3.9 as the highest. Highest

change was expected for blood pressure and heart rate. Subjects expected blood pressure to rise and heart rate to increase, when in fact they experienced a lowered blood pressure and decreased heart beat. Their expectancy of change for the computer test variables were close to 3 = no change. Since we also asked the control group

Table 3

Results of analysis of covariance of post-test means with pre-test means as covariates for main variables

Variable	Condition	SS	d.f.	<i>F</i>	<i>P</i>
Heart rate	Factor 1	74.13	3	0.974	0.41
	Factor 2	7.18	1	0.283	0.60
	Factor 1 × factor 2	142.78	3	1.876	0.14
	Error	3806.02	150		
Blood pressure (systolic)	Factor 1	85.44	3	0.643	0.59
	Factor 2	23.32	1	0.527	0.47
	Factor 1 × factor 2	191.36	3	1.441	0.23
	Error	6595.64	149		
Blood pressure (diastolic)	Factor 1	275.33	3	2.989	0.03
	Factor 2	19.06	1	0.621	0.43
	Factor 1 × factor 2	26.37	3	0.1286	0.84
	Error	4574.60	149		
Well-being	Factor 1	51.68	3	0.237	0.87
	Factor 2	125.00	1	1.722	0.19
	Factor 1 × factor 2	571.17	3	2.622	0.05
	Error	10527.62	145		

Factor 1: replication vs. suggestion; factor 2: Subject expectancy (caffeine, no caffeine, double blind, control).

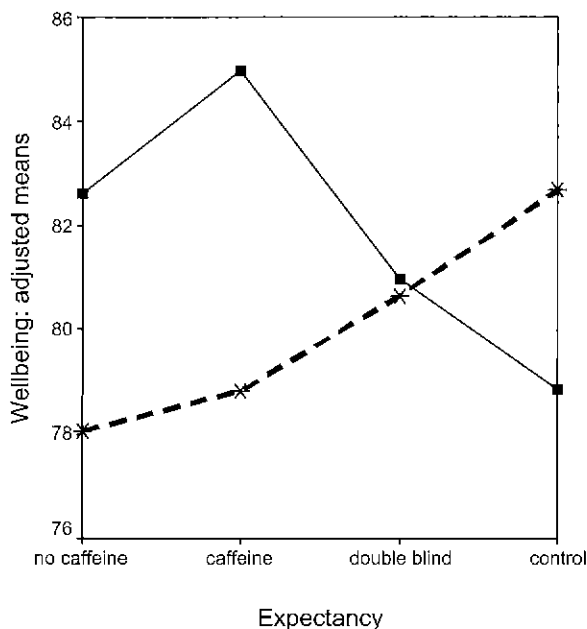


Fig. 1. Interaction plot: adjusted means for well-being by expectancy-group and replication vs. suggestion condition. Solid line: suggestion; hatched line: replication.

about what they would expect, had they been randomized into the caffeine group, we could thereby assess the social stereotype. In this group, ratings ranged from close to indifference for the computer variables to 3.8 for blood pressure. Thus expectancy for change following the consumption of a cup of coffee was rather low in the first place. We checked, whether expectancy was different for the experimental groups by exploratory Kruskal–Wallis analyses. It was only different for

blood-pressure ($P = 0.0001$), due to a significant difference between the groups 'told no caffeine' and 'told caffeine', between groups 'told no caffeine' and the control group, and between the groups 'told caffeine' and the double-blind group. There was no difference in expectancies in any of the variables in the suggestion vs. replication factor.

We analyzed whether the experimental manipulation was successful, by asking subjects after the experiments, how they rated the subjective probability of having received coffee and by subjecting these data to a one-way ANOVA. This resulted in a highly significant ($P < 0.0001$) difference, with all contrasts being highly significant. The group 'told no caffeine' rated this probability as 3.9, group 'told caffeine' as 7.1 and group 'double-blind' as 5.7 on a 10-cm visual analogue scale with anchors at '0 = surely without caffeine' and '10 = surely caffeine'.

In the debriefing interview, subjects rated the experimenters favorably and felt mostly not disturbed by the video camera. Sixty-eight percent of the subjects said they were apprehensive at the beginning of the experiment, while after the pre-measurements only 23% were still apprehensive. Eighty-four percent of the subjects rated the experiment as plausible and believable, 16% said it was somewhat believable. Nobody rated it as unbelievable or very unbelievable. Eleven percent indicated that they had second-guessed the deception. Thus, on the whole the experimental manipulation was successful instilling expecta-

Table 4

Correlations (Spearman's rho) between specific expectations of the effects of caffeine on variables with actual changes in these variables by experimental manipulations

	Heart rate	Blood pressure		Misprint correction ^a	Computer test 'Click' ^a
		(systolic)	(diastolic)		
No caffeine	-0.02	-0.22	-0.17	-0.29	0.11
Caffeine	-0.31	0.18	-0.10	-0.05	-0.02
Double-blind	-0.17	0.32*	0.35*	-0.04	0.53**
Replication	0.04	0.05	0.09	0.02	0.09
Suggestion	-0.07	0.11	0.15	-0.13	0.10

* $P < 0.05$; ** $P < 0.01$.

^a Positive values reflect better performance.

tions of coffee or no coffee, and subjects believed, what they were told.

4. Discussion

We designed this experiment, in order to find out, why an earlier experiment using the same experimental paradigm was unable to reproduce effects reported in the literature (Fillmore and Vogel-Sprott, 1992; Fillmore et al., 1994; Kirsch and Weixel, 1988), and to reproduce our own results (Walach et al., 2001). The results of this experiment are somewhat more in line with the literature and differ from our own earlier results, in two important respects. The first experiment had found one main effect on the instruction factor with the group 'told no caffeine' reporting significantly better well-being. Apart from that, there was a main effect of experimenters, which very likely was not due to experimental manipulation but personal differences, and a weak interaction towards the expected direction. The problem with this first experiment was that the significant main effect was counterintuitive and contrary to the literature. We could not come up with a ready explanation for this effect. We therefore took great care to reduce error variance as much as possible. As in the first experiment, we had experimenters totally blind to the rationale of the study, who were in charge of contact with subjects, taking data, administering experimental manipulation and giving the appropriate information. Thus, nobody of the research team who knew about the hypothesis had other than cursory contact with subjects. We deliberately displaced the experiment out of the campus to the local police academy, which is completely unconnected with the university, in order to avoid any leakage of information from the previous experiment, which had taken place 1 year earlier. We did not employ psychology students who might be knowledgeable about the research paradigm, but mostly members of the police academy, their relatives and friends, as well as some students of completely different disciplines who were personally known to the experimenters. Note that, for all we can judge from reports, our studies were the

only ones which had totally blind experimenters installed between researchers and subjects. While the standard study in the literature had experimenters who were blind to the actual group allocation of subjects, such as to prevent biased measurements and non-verbal signals, the principal investigator, who knew about purpose and hypotheses, was part of the research team in all of the studies we know of, either administering experimental manipulations, or taking measurements. We hired and paid experimenters, who liked doing their job, but were rather indifferent to outcomes, since they did not have hypotheses in the first place.

From this experiment a clear picture emerges: there is a weak instruction effect in one out of four main variables. The size of this effect can be estimated to be $r = 0.14$ or $d = 0.28$ according to Rosenthal (1991). It can be seen in the smaller reduction in diastolic blood pressure in the 'told caffeine' group, compared with the other groups. In the whole group blood pressure fell slightly from pre- to post-measurement. This seems to reflect decreased apprehension and arousal, which was present in nearly three quarters of subjects at pre-measurement. Against this general trend of decreasing arousal the expectation of caffeine seems to have worked, thus reducing this decrease significantly. However, it should not be overlooked that all the other variables — systolic blood pressure, well-being, heart rate, performance at computer — did not show clear-cut effects. Although a trend is visible with the same tendency in all variables, this tendency is so weak that it hardly ever even approaches significance. Therefore we can conclude that instruction effects of one strong cup of placebo-coffee are visible, but small.

An additional information about scientifically proven pharmacological effects of caffeine given to one-half of the subjects before drinking the coffee did not have an additional effect. There is only one interaction for the variable well-being approaching significance pointing in the hypothesized direction: subjects told to receive caffeine in the group who was given the suggestion that caffeine did indeed have pharmacological effects reported slightly improved well-being. This effect

is small and not visible in the objective variables and could thus be due to social desirability. One could argue that a simple written information is too weak as suggestion, since we know from studies that only suggestions which are accompanied by slight alterations of the bodily experiences as indicator of an upcoming effect are really effective (see Kirsch et al., 1999; Wickless and Kirsch, 1989 for more details). While this is true, all we wanted to achieve with this second experimental factor was to activate the social stereotype, if present, and not more.

We repeatedly did not succeed to reproduce instruction effects of caffeine placebos reported in the literature, at least not effects of the same magnitude. The reasons for that, in our opinion, are the following: firstly, by having blind experimenters conduct the experiment, all effects of experimenter expectancies were blocked. Effects of biased experimenters, who implicitly arouse subjects and thus alert them to watch out for hidden cognitive cues about what is expected from them, enhanced by the apprehension to 'be a good subject' in order to receive course credits (Finkelstein, 1976) — these possible distortions were absent in our design. Rosenthal's own guess was that having completely blind experimenters between biased experimenters and subjects in his own picture rating tasks would have prevented experimenter effects, although he regretted never to have done an experiment like that (Rosenthal, 1984).

The second reason is the lack of expectancies aroused by the experimental manipulations. We checked specific expectancies after the consumption of the beverage, but before post-measurements, for each variable separately. We found rather little variation. Expectancies were in the more neutral range between 3 ('no change expected') and 3.9 (4: 'slight increase/improvement expected'). In accordance with the expectancy hypothesis of placebo effects (Kirsch, 1997), expectancy was highest for blood pressure, where indeed we found the only effect. Since we also asked subjects of the zero-control group about their expectancies of what would have happened had they drank a cup of strong coffee, we also

have indications about the social stereotype of what it means to people hereabouts to drink one strong cup of coffee. Expectancies were similar and rather low in the control group, too. Therefore, we can conclude that people in Germany who are used to drinking coffee — remember that two-thirds of the subjects were used to drinking more than one cup of coffee — do not expect dramatic physiological or psychological changes from drinking one cup of coffee.

It is interesting to note that expectancies and actual changes are rather uncorrelated: there is only one significant and small correlation, which might well be a chance fluctuation considering the multiplicity of tests. Moreover, the correlation is only visible in a secondary variable — computer test 'Click' — which did not show a significant main instruction effect in the first place. Furthermore, also the group 'told no caffeine' showed a significantly smaller reduction of blood pressure than the control group. If expectancies were the main causal factors for producing changes, and not conditioned responses — like associations between taste and physiological effects — as response expectancy theory postulates (Kirsch, 1985, 1997), we should have seen a significant difference between the group 'told caffeine' and the control group, but not between the group 'told no caffeine' and the control group (see Section 2.5, hypothesis 2). Since we also found a significant contrast between the latter groups, instruction effects cannot solely be due to expectancies but do also have a conditioned part to them.

The habit of drinking coffee was hardly correlated with the results. Subjects who normally drank less coffee had a slightly stronger effect in systolic blood pressure, and thus were marginally more sensitive ($r = 0.2$, $P < 0.02$). All other variables were uncorrelated. Thus, our results are not confounded by coffee consumption habits.

We double-checked the success of our experimental manipulation. On the whole, it was successful. The subjective probability of having received caffeine varied substantially and significantly between experimental groups according to their experimental information. However, it should not be overlooked that experimental ma-

nipulations were not totally successful, as some subjects in the ‘told no caffeine’ group still believed to receive coffee, and some subjects in the ‘told caffeine’ group did not believe this. An analysis with those subjects excluded yielded the same results. Thus, while overall we achieved successful manipulations, some error variance is introduced by a few subjects with expectations contrary to the experimental assignment. In an experimental paradigm with such small effects this reduces power even more. Yet this does not invalidate our results, since in nearly all expectancy studies manipulations are far from perfect. We used twice as many subjects than the studies reported in the literature. Our study had a power of 77% to detect a medium sized effect of $f = 0.25$.

In conclusion, then, we could demonstrate only a weak effect of caffeine instruction and consumption of decaffeinated coffee on diastolic blood pressure: the reduction of blood pressure seen in the control group was significantly contravened in the groups ‘told caffeine’ and ‘told no caffeine’. Heart rate, systolic blood pressure, well-being, and performance at two cognitive tasks was not influenced by experimental instructions. Information that coffee had proven pharmacological effects did not show any effect. Expectancies for change after drinking one cup of coffee were generally very low, and only blood pressure was expected to change slightly. Actual changes and expectancies for change were uncorrelated. Apart from the weak social stereotype of the effects of coffee it could be the tightly controlled design of our experiment, with completely blind experimenters and subjects, which was responsible for the failure to reproduce effects known from the literature. We therefore suggest that effects reported in the literature are reproduced with tighter controls, such as blinded experimenters, before they are taken for granted.

Acknowledgements

This study was funded by the Institut für Grenzgebiete der Psychologie und Psychohygiene,

Freiburg, Germany. We are grateful to Dr Joachim Kepplinger and LPD Dieter Huser of the police academy Freiburg (Landespolizeischule Baden-Württemberg), who made it possible for us to conduct the study at this venue and who supported us in every respect. Frank Lenz programmed the computer test ‘Click’ for us. Professor J. Fahrenberg advised us as to blood pressure measurements.

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. *Pharmacol. Biochem. Behav.* 40, 115–119.
- 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.
- Finkelstein, J.C., 1976. Experimenter expectancy effects. *J. Commun.* 26, 31–38.
- Hedges, L.V., Olkin, I., 1985. *Statistical Methods for Meta-Analysis*. Academic Press, Orlando, FL.
- Hobi, V., 1985. Basler Befindlichkeits-Skala. Ein Self-Rating zur Verlaufsmessung der Befindlichkeit. Beltz Test Gesellschaft, Weinheim.
- 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. *Psychol. Bull.* 99, 347–360.
- Kirsch, I., 1985. Response expectancy as a determinant of experience and behavior. *Am. Psychol.* 40, 1189–1202.
- Kirsch, I., 1997. Specifying nonspecifics: Psychological mechanisms of placebo effects. In: Harrington, A. (Ed.), *The Placebo Effect: Interdisciplinary Explorations*. Harvard University Press, Cambridge, MA, pp. 166–186.
- 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. *Behav. Neurosci.* 102, 319–323.
- Kirsch, I., Wickless, C., Moffit, K.H., 1999. Expectancy and suggestibility: are the effects of environmental enhance-

- ment due to detection? *Int. J. Clin. Exp. Hypnosis* 47, 40–45.
- Lienert, G.A., 1955. Die Bedeutung der Suggestion in pharmakopsychologischen Untersuchungen. *Zeitschr. Exp. Angew. Psychol.* 3, 418–438.
- Marlatt, G.A., Rohsenow, D.J., 1980. Cognitive processes in alcohol use: expectancy and the balanced placebo design. In: Mello, N. (Ed.), *Advances in Substance Abuse*. JAI Press, Greenwich, pp. 159–199.
- Murray, J.B., 1988. Psychophysiological aspects of caffeine consumption. *Psychol. Rep.* 62, 575–587.
- Rosenthal, R., 1984. Interpersonal expectancy effects and psi: some communalities and differences. *New Ideas Psychol.* 2, 47–50.
- Rosenthal, R., 1991. *Meta-Analytic Procedures for Social Research*. Sage, Newbury Park.
- Schweizer, K., 1989. Eine Analyse der Konzepte, Bedingungen und Zielsetzungen von Replikationen. *Archiv Psychol.* 141, 85–97.
- Walach, H., Schmidt, S., Wiesch, S., Bühr, N., 2001. The effects of a caffeine placebo and experimenter expectation on blood pressure, heart rate, well-being and cognitive performance: a failure to reproduce. *Eur. J. Psychol.* 6(1), 15–25.
- Westhoff, G., 1993. *Handbuch psychosozialer Messinstrumente. Ein Kompendium für epidemiologische und klinische Forschung zu chronischer Krankheit. [Handbook of Psychosocial Measurement Instruments. A Compendium for Epidemiological and Clinical Research on Chronic Diseases]*. Hogrefe, Göttingen.
- Wickless, C., Kirsch, I., 1989. Effects of verbal and experiential expectancy manipulations on hypnotic susceptibility. *J. Person. Soc. Psychol.* 57, 762–768.
- Zwyghuizen-Doorenbos, A., Roehrs, T.A., Lipschutz, L., Timms, V., Roth, T., 1990. Effects of caffeine on alertness. *Psychopharmacology* 100, 36–39.