HUMAN PSYCHOPHARMACOLOGY
Hum. Psychopharmacol Clin Exp 2009; **24**: 549–558.
Published online 20 August 2009 in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/hup.1054

Does the presence of a pharmacological substance alter the placebo effect?—results of two experimental studies using the placebo-caffeine paradigm†

Harald Walach[1]* and Rainer Schneider[2]

[1]*School of Social Sciences, European Office of the Samueli Institute for Information Biology, University of Northampton, UK*
[2]*Department of Human Sciences, Differential Psychology and Personality Psychology, University of Osnabrück, Osnabrück, Germany*

Objectives We employed the placebo-caffeine paradigm to test whether the presence or absence of a substance (caffeine) influences the placebo effect. Methods: In experiment 1 consisting of four conditions with $n = 15$ participants each (control, placebo, two double-blind groups, each with placebo only), we maximized the placebo effect through expectation. Effects were assessed with physiological (blood pressure, heart rate), psychomotor (response times), and well-being indicators (self-report). In experiment 2, caffeine was administered in one of the double-blind groups, and another condition was added where caffeine was given openly.
Results Effect sizes were medium to large for some outcome parameters in experiment 1 and 2, showing partial replicability of the classical placebo effect. Although not formally significant, differences between the double blind placebo conditions of the two experiments (with and without caffeine present) were medium to small. There was a significant difference ($p = 0.03$) between experiment 1 and experiment 2 in the physiological variables, and a near significant interaction effect between groups and experiments in the physiological variables ($p = 0.06$).
Conclusion The question warrants further scrutiny. The presence of a pharmacological substance might change the magnitude of the placebo response. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS — placebo effect; replication; caffeine; blood pressure; psychomotor performance; double blind

INTRODUCTION

Since their introduction in 1948, randomized placebo controlled clinical trials are the cornerstone of pharmacological efficacy testing. Placebo, i.e., a substance devoid of active agents, is used to differentiate specific pharmacological effects from artificial ones like regression to the mean, natural course of disease, or spontaneous fluctuations, as well as non-specific psychological factors of hope, alleviation of anxiety, and expectancy (Hróbjartsson, 2002). The unpredictability, variability, and strength of placebo responses in clinical trials may compromise their interpretation (Kirsch *et al*., 2008). Hence, research into the factors that determine the magnitude of responses to placebo and treatment is important.

Placebo responses are partially a function of expectations induced in participants by the ritual of a study (Moerman, 2002; Moerman and Jonas, 2002). A meta-analysis found an effect size of $d = 0.15$ for responses in placebo groups of clinical pain trials versus no-treatment controls, but an effect size of $d = 0.93$ for placebo response versus no treatment for experimental studies, which aim at maximizing placebo effects (Vase *et al*., 2002). This shows that the placebo effect can be large to a clinically meaningful degree when enhanced by expectation, personal contact, and ritual. A study comparing different levels of personal caring during the application of a placebo-acupuncture condition could make plausible that the level of involvement of a practitioner during a trial is one crucial variable, likely enhancing expectations (Kaptchuk *et al*., 2008). A series of studies using neuropharmacological probes have made it clear that part of the clinical placebo response is mediated by expectancy and functional neurotransmitter systems involved in the processing of expectancy, pleasure, and pain (Amanzio *et al*., 2001; Benedetti, 1996; Colloca and Benedetti, 2005; Enck

* Correspondence to: Professor H. Walach, University of Northampton, School of Social Sciences & Samueli Institute for Information Biology, Boughton Green Rd, Northampton NN2 7AL, UK. Tel: 0044 1604 89 2952. Fax: 0044 1604 722067. E-mail: harald.walach@northampton.ac.uk
† The research was conducted at University Hospital, Institute of Environmental Medicine and Hospital Epidemiology, Freiburg, Germany.

Received 26 October 2008
Accepted 3 July 2009

et al., 2008; Pollo *et al*., 2001, 2002), while learning and conditioning may play another important role (Amanzio and Benedetti, 1999; Benedetti *et al*., 1998; Colloca *et al*., 2008a, b). Just how powerful such placebo effects in clinical trials can be was demonstrated impressively by a series of large, three-armed studies comparing sham-acupuncture and real acupuncture with conventional treatment in various pain syndromes (Haake *et al*., 2007; Scharf *et al*., 2006). They found that, while sham and real acupuncture were indistinguishable, they both were twice as powerful as conventional treatments, including pharmacotherapy, mobilization, and physiotherapy. This finding has just been replicated (Cherkin *et al*., 2009).

Such results create a paradox (Walach, 2001): They show that a placebo treatment, presumably operating via the expectancy or conditioning history of patients can be much more powerful than supposedly evidence-based treatments that have been tested against placebo controls themselves. Clearly, these effects created by the application of allegedly inert treatments are highly variable.

If this variability were a problem of individual trials only, then meta-analytic findings of correlations between treatment responses and placebo responses in clinical trials should reveal non-significant correlations under the condition that the intervention is substantial and the placebo intervention, by and large, negligible. When formally tested this is not the case, but significant correlations of varying magnitudes between improvement rates in treatment and placebo groups of clinical trials have been found. In an analysis of 26 RCTs with treatment duration of more than 12 weeks, a significant correlation of $r = 0.59$ between response rates of placebo and treatment groups was reported (Walach and Maidhof, 1999). In a similar analysis of antidepressants, Kirsch and Sapirstein (1998) found a very high correlation of $r = 0.90$. In another analysis of data of six most widely prescribed antidepressants submitted to the FDA for approval, Kirsch *et al*. (2002) showed that about 80% of the response to medication was duplicated in placebo control groups. An analysis of 141 long-term trials with a more heterogeneous set of studies yielded a highly significant correlation between placebo and medication response rates across different diseases ($r = 0.78$) (Walach *et al*., 2005). Most importantly, placebo response rates in clinical trials were only partially due to methodological artifacts and only partially dependent on the diagnoses treated. An independent meta-analysis of the placebo effect in irritable bowel syndrome also found a significant correlation between placebo and treatment response rates of $r = 0.36$, which was independent of study quality (Patel *et al*., 2005)

Taken together, these findings suggest that placebo response rates in clinical trials, across diagnoses, designs and treatment duration are sizeable and correspond in magnitude to treatment response rates. If expectancy is the major driver of placebo response rates in clinical trials, then this would mean that expectancies induced in patients during trials vary systematically with the magnitude of the effect expected for treatment. Another way of understanding this effect stems from a new systems perspective. A theoretical model predicts correlations in sufficiently closed systems between elements of the system that are not due to local mechanistic interactions (Atmanspacher *et al*., 2002; Stillfried and Walach, 2006). It would predict correlations between placebo response rates and treatment response rates in clinical trials. Such correlations would be due to formal, systemic reasons alone, although the model is not yet precise enough to allow for further quantitative predictions. Put differently, the model suggests that part of the placebo effect in clinical trials is due to a non-local correlation of placebo groups with treatment groups. We call such an effect ''non-classical placebo effect'' to distinguish it from its classical counterpart that is due to psychological or learning effects, and, in trials, to various artifacts.

We used an experimental approach to test this prediction. We reasoned that, all things being equal, the placebo effect in a study where only placebo and suggestion is used should be different from a study, where an active pharmacological agent is employed.

We therefore conducted an experimental trial that aimed at producing a purely psychological (or classical) placebo effect by inducing expectations, but without an active substance. In a second experiment, simulating a clinical trial, an active ingredient was used in addition. These experiments were deliberately designed and run as independent trials. Collapsing the outcomes would merely serve as a means of testing the theoretical assumption that non-local placebo effects could occur. We employed the placebo-caffeine paradigm for which placebo effects have been found (Anderson and Horne, 2008; Andrews *et al*., 1998; Fillmore and Vogel-Sprott, 1992; Fillmore *et al*., 1994; Fillmore *et al*., 2002; Flaten and Blumenthal, 1999; Lotshaw *et al*., 1996; Malani *et al*., 2008; Mikalsen *et al*., 2001; Schneider *et al*., 2006), although not always (Walach *et al*., 2001, 2002). We attempted to induce a placebo effect by verbal and written information about the known effects of caffeine. We conducted two experiments: one in which

DOI: 10.1002/hup

no active pharmacological ingredient was used at all, although participants and experimenters were under the impression that caffeine would be introduced in a double blind condition, where in fact only placebo caffeine was administered. The second experiment was an exact replication of the first, with the exception that an active ingredient, caffeine, was introduced in the double-blind condition.

MATERIALS AND METHOD

Study design

We conducted two experiments sequentially. For theoretical reasons these experiments were organizationally independent, participants were not randomized to experiments, but only to conditions within experiments, and experimenters were under the impression that the second experiment was a replication of the first. In the first experiment only placebo was used, and we randomized participants into four groups ($n = 15$ in each group): one group received no beverage and was only measured as a control condition for random fluctuation, one group received a placebo with the information that caffeine was being administered and two groups received placebo under double-blind conditions (double blind X and Y). In the second experiment, we administered caffeine in the double blind condition Y, mimicking a clinical trial, and we added a group that received caffeine and was told so (open caffeine; $n = 15$). All procedures and participant contacts were handled by the same two female experimenters, who were blind to the actual purpose of the study and the substances administered.

Experiment 1

Procedures. Participants responded to newspaper advertisements announcing a study investigating the effects of caffeine on well being, arousal, and cognitive performance. Participants were deemed eligible for the study if they had no medical impairment and agreed to abstain from caffeine consumption for 24 h and to fast for 4 h prior to participation. To test for compliance, random saliva specimens (five per condition totaling 20) were taken and screened for etofyllin. Samples were stored at -80°C and analyzed using an enzyme-immunoassay (EIA) method developed in the Department of Forensic Medicine of the University Hospital Freiburg. Samples were considered satisfactory if the saliva concentration was below 500 ng/ml caffeine. The study was carried out at the Institute for Environmental Medicine and Hospital Epidemiology Freiburg, Germany. Ethical approval for the experimental procedures was obtained from the Hospital's Ethics Committee. Randomization to the experimental groups was done with the software RITA (Pahlke *et al*., 2004). We used a quinine hydrochloride solution (0.1%) as placebo to match the bitter taste of pure caffeine in a concentration of 1 ml/10 kg body weight, sampled with a syringe and mixed with 100 ml orange juice, following effective published procedures (Flaten and Blumenthal, 1999). In the ''double-blind condition,'' participants were given the information that the liquid contained caffeine with a 50% probability. In the placebo condition, participants were told they would receive caffeine.

Sessions were scheduled in the morning (8:00–12:00). At the beginning of the sessions, participants were given written information about the alleged purpose of the study and the physiological, mood, and performance enhancing effects of caffeine. They then gave their written informed consent. Each session consisted of a baseline assessment of well being, cardiovascular measures, and cognitive performance followed by experimental intervention and post-treatment assessments. Physiological measures were taken three times with 2 min rest intervals and averaged. Then, baseline values for well being were taken. Finally, baseline measures for reaction time (RT) were taken after participants achieved an accuracy rate of 80% in a familiarization test.

Following the baseline measurements, the experimenter opened a sealed and numbered envelope containing the random assignment to the experimental group. Participants were asked to rate how they expected the beverage would affect their well being, physiology, and cognitive performance. They were then weighed and given the appropriate amount of ''caffeine'' in juice. This had to be consumed within 1 min. This was followed by a waiting period of 30 min ''to allow the caffeine to take effect,'' during which participants rested and read magazines. After the waiting period, they were asked to rate how the beverage affected them and how certain they were that they had consumed caffeine. Then, post-treatment measures were taken in the same order as previously. Participants were remunerated with €15.

Participants. Sixty participants, 14 males and 46 females with an average age of 32.3 (SD = 11.9; range = 20–64 years) took part in experiment 1. All were normotensive, not taking any prescription medication and were not suffering from any medical condition, not receiving treatment for problems with alcohol and/or drug use, and not breast-feeding. They

DOI: 10.1002/hup

reported to consume on a daily basis on average 1.6 cups of regular coffee and 1.3 cups of tea.

Measures

Blood pressure. Blood pressure (systolic and diastolic) and heart rate were measured with a calibrated digital oscillometric sphygmomanometer, the OMRON M5 Professional (Omron Inc, Germany) which automatically inflates the arm cuff and shows and stores the values on a LCD display. Participants sat on a chair relaxing with their extended left arm lying on a cushion. The cuff was wrapped around the upper arm, with the lower edge placed 1–2 cm above the inner side of the elbow joint. The level of the cuff was placed at the same level as the heart during measurement.

Well being. A 24-item, three-dimensional validated German questionnaire (Multidimensional Well-being Questionnaire) was administered (Steyer *et al*., 1997). It assesses well being according to the dimensions alertness/weariness, positive/negative mood, and calmness/disconcertment, and has been widely used in German psychopharmacological studies. Each scale consists of eight bipolar items with five anchors from which minima and maxima are labeled (not at all—very much so). Internal consistency and test–retest reliability of all scales is very good ($\geq$0.87).

Simple reaction time task. For 200 consecutive trials, participants were required to make quick key press responses to visual stimuli. The task consisted of four letter-color combinations (X, Y, red, blue) which were to be discriminated and responded to by pressing one of two adjacent keys on the computer keyboard using the index and middle fingers of the right hand. The four stimulus combinations were randomly balanced across each test session. The letter X in red and the letter Y in blue were to be responded with key 1. The letter X in blue and the letter Y in red were to be responded with key 2. Each stimulus was displayed for a maximum of 2000 ms and the computer screen was blank for a varying interstimulus interval of 1000–2000 ms. After each response, feedback was provided in the upper left hand corner of the screen (correct, false). Response execution was measured by the mean RT to the stimuli. Reactions equal or larger than 90 ms were deemed valid. A test was completed in approximately 12–15 min. The task was designed by RS and operated by E-Prime (Psychology Software Tools, Pittsburgh, PA) (Schneider *et al*., 2002).

Experiment 2

Study design and procedure. All aspects of experiment 2 were the same as in experiment 1. In contrast to experiment 1, however, two major design alterations were implemented. In the double blind condition Y, participants received 3 mg caffeine/kg per body weight but were told that the probability to be given caffeine was 50%. In an additional experimental condition, participants were administered a dose of 3 mg caffeine/kg body weight and informed about this (open caffeine condition). The caffeine solution (3%) was given in amounts of 1 ml/10 kg body weight, sampled with a syringe, and mixed in 100 ml orange juice. The two female experimenters, who were identical to experiment 1, were told that this additional condition investigated the effects of caffeine in a different carrier solution (note that in the double blind condition of experiment 1 experimenters believed to administer caffeine). Post study assessment of the plausibility of this experimental condition showed that the cover story did not arouse suspicion and that the blinding of the experimenters was uncompromised.

Participants. Seventy five participants, 25 males and 50 females with an average age of 29.9 (SD = 10.3; range = 19–60 years) participated in experiment 2. Inclusion and exclusion criteria were the same as in experiment 1. Participants reported to consume on a daily basis on average 1.3 cups of regular coffee and 1.2 cups of tea.

Measures. The same measures were applied as in experiment 1.

Data analysis. Analyses were performed using the statistical packages SPSS 14.0 and Statistica. Power calculations were made based on findings from two recently conducted experiments with the same paradigm (Schneider *et al*., 2006). To determine treatment effects, ANCOVAs were carried out with the baseline measures as covariates. Simple contrasts were applied where appropriate. Effect sizes were calculated according to Cohen (1988). In experiment 1, for all dependent variables, significant differences between the placebo group and the control group (placebo effect) and the double-blind groups X and Y (placebo response) were expected. In experiment 2, a significant difference between the placebo group and the control group was expected (placebo effect). Furthermore, significant differences were also expected between the double blind group Y and the control group (pharmacological effect) and the caffeine group and the control

DOI: 10.1002/hup

group (pharmacological effect + expectancy). To see whether the presence of a substance in double-blind arms makes a difference, a randomly selected group of the double-blind arm of experiment 1 (i.e., double-blind placebo) was tested against the double-blind placebo group of experiment 2 (non-classical placebo effect). Expectancy measures were correlated with difference scores of appropriate outcome variables using the Spearman rank correlation coefficient. To test for significance, *p*-levels were adjusted according to Holm (1979). Specifically, for the physiological parameters systolic and diastolic blood pressure, and heart rate, the significance level was set at $p \leq 0.017$ (0.05/3). Accordingly, for psychomotor performance the number of correct reactions and mean reaction time the significant level was set at $p \leq 0.025$ (0.05/2). Finally, for the well-being measures mood, calmness, and alertness the significance level was set at $p \leq 0.017$ (0.05/3). For none of the measures it was specified which one would have to be significant. This was in line with our reasoning that placebo effects may show in varied, to some extent unpredictable ways.

RESULTS

Experiment 1

Physiological measures. All measures (Table 1) were normally distributed and fell within the range of normal physiological values. Re-test reliabilities of the three measurements before and after treatment were high and showed no outliers which could have affected measurement validity (baseline: systolic blood pressure $r = 0.91$; diastolic blood pressure $r = 0.85$; heart rate $r = 0.93$; post-treatment measures: $r = 0.92$; $r = 0.82$; $r = 0.91$, respectively).

The ANCOVA for systolic or diastolic blood pressure failed to yield significant differences between the four groups [systolic: $F(3, 55) = 0.99$; $p = 0.41$; diastolic: $F(3, 55) = 1.3$; $p = 0.28$]. Likewise, heart rate did not differ between the groups [$F(3, 55) = 0.42$; $p = 0.99$]. Participants of the placebo group displayed the highest values (Table 1). Calculation of the effect sizes for the differences between the placebo group and the control group yielded $d = 0.55$ for systolic blood pressure, $d = 0.06$ for diastolic blood pressure, and $d = -0.05$ for heart rate (see Table 2).

Reaction time. Analyses did not reveal any differences between the four groups neither with regard to the mean number of correct responses [$F(3, 55) = 0.7$; $p = 0.56$] nor the average reaction time [$F(3, 55) = 0.63$; $p = 0.60$]. Hence, hypothesis 1 assuming a placebo effect could not be confirmed. The effect sizes for the number of correct responses and mean reaction time were $d = 0.25$ and $d = 0.34$.

Subjective well being. For neither of the three well being dimensions, a significant difference between the placebo group and the control group was found (calmness: $F(3, 55) = 0.49$; $p = 0.69$, mood: $F(3, 55) = 1.11$; $p = 0.35$, and alertness: $F(3, 55) = 1.92$; $p = 0.14$). Effect sizes (Table 2), however, showed that participants reported mild to strong mood changes corresponding to the alleged effect of the beverage (calmness: $d = -0.33$, alertness: $d = 0.66$, and mood: $d = 0.43$).

Table 1. Mean and standard deviation (SD) for baseline and post-treatment measures of blood pressure (mmHg), heart rate (beats per minute), correct responses, reaction time (ms), mood, alertness, and calmness in experiment 1

	Experimental group							
	Placebo		Double blind X		Double blind Y		Control	
	Baseline	Post	Baseline	Post	Baseline	Post	Baseline	Post
SBP[a]	116.6 (10.2)	116.4 (10.6)	114.9 (14.5)	114.9 (13.5)	111 (12.5)	111.8 (9.0)	112.4 (12.8)	109.8 (13.2)
DBP[b]	70 (4)	69.5 (4.9)	71.2 (8.7)	71.8 (9)	69.6 (10.2)	71.7 (9.1)	70.6 (10.4)	69.8 (8.7)
HR[c]	71.4 (14.5)	66.7 (13.5)	66.5 (12.6)	62.8 (10.7)	74.2 (12.2)	69.6 (12.5)	72.3 (12.5)	67.7 (10.3)
Hits[d]	184.9 (14.1)	194.2 (5.8)	185.7 (11.3)	191.8 (7.4)	178.1 (19.8)	190.3 (9.4)	189.7 (8.9)	194.2 (6)
RT[e]	830 (174)	761 (149)	850 (156)	768 (138)	904 (148)	827 (117)	859 (167)	759 (117)
Mood[f]	30.9 (6.5)	32.1 (5.6)	33.6 (3.6)	34.5 (3)	33.5 (4.1)	33.3 (3.8)	29.9 (4.8)	30.3 (5)
Alertness[f]	27.4 (7.5)	29.1 (6.4)	29.1 (7.7)	30.2 (5.9)	29.1 (8.2)	27.6 (5.2)	26.2 (7.1)	25.4 (7.1)
Calmness[f]	30.5 (5.5)	30.9 (4.4)	32.5 (4)	32.7 (3.3)	32.3 (6.2)	31.6 (7.3)	31.3 (4.3)	32.7 (4.2)

[a]Systolic blood pressure.
[b]Diastolic blood pressure.
[c]Heart rate.
[d]Number of correct responses.
[e]Reaction time.
[f]Range: 8–40; high value indicates good mood, high alertness, high calmness.

DOI: 10.1002/hup

Table 2. Effect sizes d for the placebo effect and differences between placebo groups of double blind arms of experiments 1 and 2

	Experiment 1	Experiment 2	
	Classical placebo effect[a]	Classical placebo effect	Non-classical placebo effect[b]
SBP[c]	0.55	−0.05	0.58
DBP[d]	0.06	0.22	0.07
HR[e]	−0.05	0.03	−0.13
Hits[f]	0.43	0.12	0.22
RT[g]	0.25	−0.58	0.25
Mood	0.43	0.31	−0.06
Alertness	0.66	0.64	−0.31
Calmness	−0.33	−0.43	0.16

[a]Classical placebo effect (placebo vs. control).
[b]Non-classical placebo effect (double blind placebo experiment 1 vs. double blind placebo experiment 2).
[c]Systolic blood pressure.
[d]Diastolic blood pressure.
[e]Heart rate.
[f]Number of correct responses.
[g]Reaction time.

Experiment 2

Physiological measures. As in experiment 1, all measures (Table 2) were normally distributed and re-test reliabilities of the three measurements before and after treatment were high (baseline: systolic blood pressure $r = 0.84$; diastolic blood pressure $r = 0.82$; heart rate $r = 0.94$; post-treatment measures: $r = 0.89$; $r = 0.86$; $r = 0.93$, respectively). One outlier with values of more than 2 standard deviations beyond the group mean was omitted from the analysis.

The ANCOVA for systolic blood pressure yielded a highly significant effect ($F(4, 68) = 6.73$; $p < 0.01$). Single contrasts revealed significant differences between the double blind group Y (caffeine) and the double-blind placebo group ($d = 1.43$) and the control group ($d = 1.37$), as well as between the open caffeine group and the placebo group ($d = 1.27$) and the control group ($d = 1.26$; cf. Figure 1). Whilst for the placebo hypothesis (placebo group versus control group) no effect was found ($d = -0.05$), there was a medium (but non significant) difference between a randomly selected placebo group of the double-blind arm of experiment 1 and the double-blind placebo group of this experiment of $d = 0.58$ (cf. Table 2).

For diastolic blood pressure, a significant effect [$F(4, 68) = 13.57$; $p < 0.01$] was found. Significant differences were seen between the double blind group Y (blind caffeine) and the double blind group X (blind placebo) ($d = 1.91$), the open placebo group ($d = 1.96$), and the control group ($d = 2.15$), as well as between the caffeine group and the double blind group X (blind placebo) ($d = 1.2$), the open placebo group ($d = 1.26$), and the control group ($d = 1.46$; cf. Figure 1). Again, the placebo hypothesis could not be confirmed ($d = 0.22$), and there was no difference between the placebo effect in the double blind-placebo group of experiment 2 and a randomly selected double-blind placebo group or experiment 1 ($d = 0.07$).

The ANCOVA for heart rate did not reveal a significant difference between the groups surpassing the threshold for multiple testing [$F(4, 67) = 2.68$; $p = 0.04$).

Reaction time. Analyses did not reveal any differences between groups neither with regard to the mean number of correct responses [$F(4, 69) = 1.14$; $p = 0.34$] nor for the average reaction time [$F(4, 69) = 0.52$; $p = 0.72$]. Effect sizes for the placebo effect hypothesis were $d = -0.58$ for correct responses and $d = 0.12$ for reaction time. Differences between the

Table 3. Mean and standard deviation (SD) for baseline and post-treatment measures of blood pressure (mmhg), heart rate (beats per minute), correct responses, reaction time (ms), mood, alertness, and calmness in experiment 2

	Experimental group									
	Control		Placebo		Double blind X		Double blind Y		Caffeine	
	Baseline	Post	Baseline	Post	Baseline	Post	Baseline	Post	Baseline	Post
SBP[a]	117.4 (9.1)	116.1 (11.9)	108.8 (10.1)	107.3 (8.3)	110.6 (11.3)	112.8 (12.5)	121.1 (8)	127.9 (8.5)	111.9 (12)	118.7 (13.3)
DBP	75 (7.7)	73.3 (7.3)	68.3 (8.2)	68.5 (8.2)	70.3 (6.9)	70.6 (8.2)	72.5 (7.8)	80.9 (8.3)	71.9 (9.1)	77.3 (9)
HR	81.1 (10.1)	76.2 (10.9)	73.2 (10.4)	70.2 (11.5)	76.8 (14.9)	72.2 (12.5)	69.2 (9.8)	65.6 (11)	79.9 (23.2)	70.9 (23.5)
Hits	182.3 (18.6)	194.1 (6.1)	181.5 (17.1)	190.8 (9.1)	184.7 (17.5)	193.1 (7.9)	186.3 (7.8)	193.1 (5.9)	185.5 (7.8)	194.9 (3.8)
RT	903 (180)	809 (151)	847 (161)	777 (140)	857 (201)	803 (149)	911 (160)	802 (142)	913 (139)	835 (163)
Mood	31.6 (5.2)	32.7 (4.5)	32.9 (5.1)	34.5 (3.5)	32.7 (4.8)	34.2 (5.3)	30.5 (4.7)	31.1 (3.9)	30.8 (4.8)	33.1 (4.3)
Alertness	27.5 (6.7)	26.9 (5.6)	28.3 (8.3)	30.1 (8.1)	28.3 (6.4)	28.9 (6.6)	26.5 (6.2)	27.4 (5.4)	26.1 (7.8)	31.5 (4)
Calmness	30.5 (5.3)	32.4 (5.1)	32.8 (3.6)	32.5 (3.3)	31.5 (4.7)	32.2 (4.7)	30.3 (4.5)	29.1 (3.7)	30 (5.2)	30.3 (6.3)

[a]Captions see Table 1.

DOI: 10.1002/hup

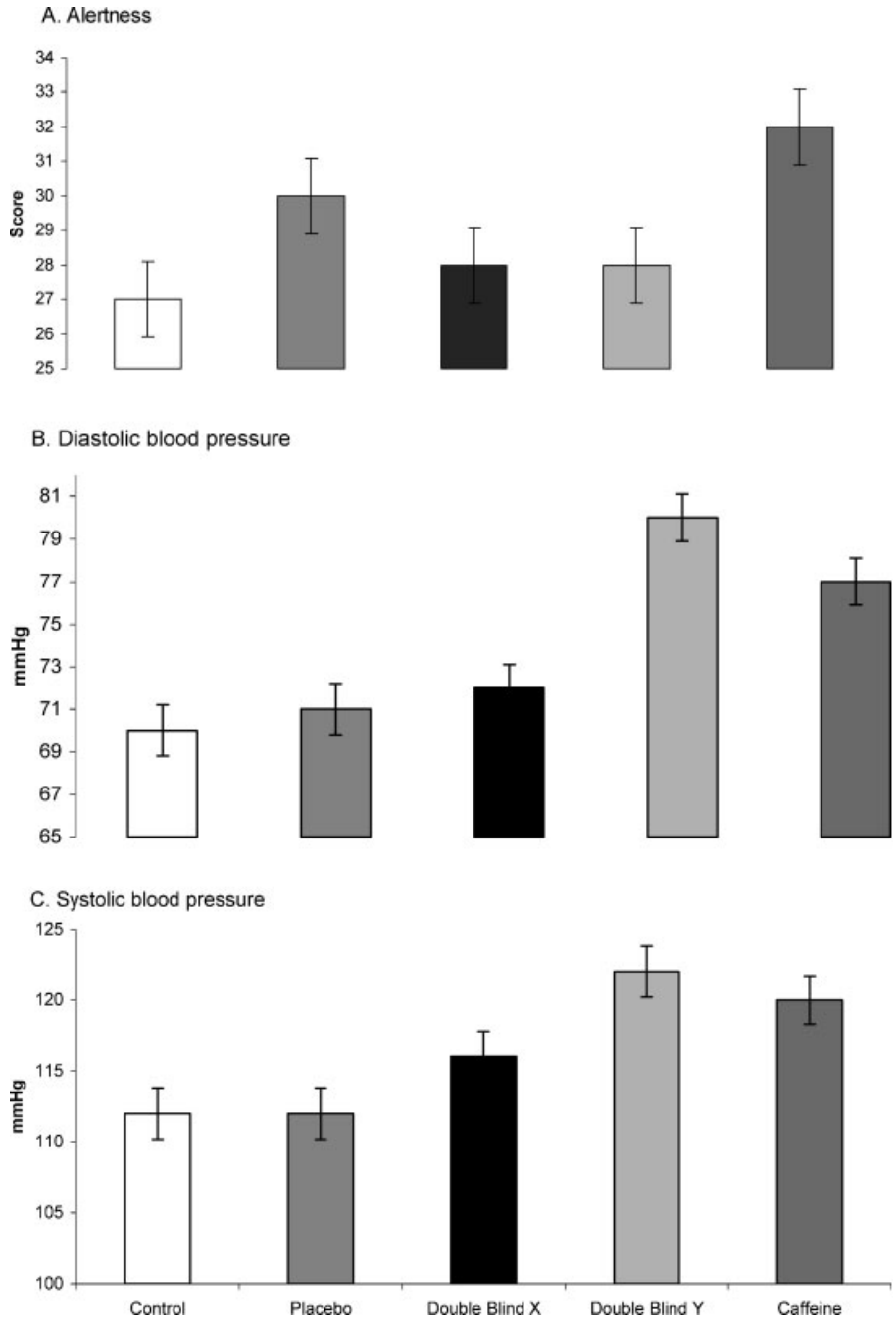

Figure 1. Means (standard error bars) for alertness, diastolic, and systolic blood pressure adjusted for baseline measures; experiment 2

placebo groups of the double blind arms of experiments 1 and 2 were $d = 0.25$ and $d = 0.22$, respectively.

Subjective well being. The analysis revealed no differences for calmness ($F(4, 69) = 0.49$; $p = 0.69$) or mood ($F(4, 69) = 1.76$; $p = 0.15$). With regard to alertness, however, a highly significant effect ($F(4, 69) = 3.67$; $p = 0.009$) was found which was due to the difference between the caffeine group and the control group ($d = 1.29$; cf. Figure 1). The effect sizes for the classical placebo hypothesis was $d = 0.64$ and the difference between double blind placebo groups (non-classical placebo effect) was a non-significant $d = -0.31$.

DOI: 10.1002/hup

Expectancies. The experimental manipulation produced highly significant differences in expectancies (Wilks' $\lambda = 0.52$; $F(12,135) = 3.19$; $p < 0.0005$) conforming to the experimental manipulation in experiment 1, i.e., highest expectancy with suggestion of caffeine, lower with double-blind suggestions but still higher than control. This difference was no longer visible when participants were asked to rate their subjective estimate whether they had experienced an effect after consumption of the beverage (Wilks' $\lambda = 0.71$; $F(12,140) = 1.65$; $p = 0.086$). Expectancies were less pronounced in experiment 2, but consonant with the experimental suggestions [Wilks' $\lambda = 0.71$; $F(16,205) = 1.52$; $p = 0.09$]. Correlations between expectancies and difference measures of outcome variables were low and inconsistent. In experiment 1, there was one significant correlation between expectancy and difference in systolic blood pressure ($\rho = 0.62$) in the double-blind group X, and one significant negative correlation between expectancy and alertness ($\rho = -0.57$) in the placebo group. Expectancies and outcome were most consistently correlated for the psychological well-being variables and performance variables in the control group ($\rho > 0.5$). There were no significant correlations between expectancies and differences in outcomes in experiment 2.

In addition to the robust *t*-tests comparing the double-blind placebo group of experiment 2 with the randomly selected double-blind group X (placebo) of experiment 1 as a test for a non-classical placebo effect, we carried out a multivariate repeated measures analysis of variance for all outcome variables between these experiments. There were no significant effects. Effect sizes are shown in Table 2.

DISCUSSION

In this study, we explored classical placebo effects, i.e., effects due to psychological processes such as expectancy, and non-classical placebo effects, i.e., effects supposedly due to the systemic set-up of a study, by simulating an experimental and a clinical trial. We hypothesized that in both experiments placebo effects due to expectancy would be visible when comparing physiological, reaction, and well-being responses in participants given placebo compared with participants of the control group. In neither experiment did we find such a placebo effect, contrary to our predecessor study (Schneider *et al*., 2006). However, the effect sizes for a number of variables were rather large indicating that the studies merely lacked statistical power. In contrast to, for instance, placebo effects found for pain analgesia and other areas (Benedetti, 2002; de la Fuente-Fernández *et al*., 2002; Levine *et al*., 1978), placebo caffeine effects appear less predictable and more varied (Walach *et al*., 2001, 2002). Given the common beliefs about the effects of caffeine on human functioning one might speculate whether placebo caffeine effects are primarily visible during phases of strain and fatigue as shown by Anderson and Horne (2008). However, we paid special attention to the fact that participants abstained from caffeine prior to the experiment. It appears that the pharmacokinetics of caffeine (i.e., what the body does to the drug) and its pharmacodynamics (i.e., what the drug does to the body) are still not understood well enough (Schneider, 2009) despite a growing body of evidence of caffeine effects in general (Chambers, 2009; Hughes, 1992; Svenningsson *et al*., 1997).

The elusiveness of the placebo caffeine effect together with its unknown size and sample characteristics points to the importance of grounding such research in effect sizes rather than significance levels alone (Vickers, 2003). The results of this study are quite sizeable but did not replicate the effect sizes of our previous study (Schneider *et al*., 2006), thus missing significance.

It is interesting to note that the effect sizes for well being (alertness, calmness, mood) in experiment 1 were replicated in experiment 2 (cf. Table 2). This is, to our knowledge, the first direct formal replication of a placebo caffeine experiment where only the participants varied. The fact that this pattern showed only for well being aligns with our previous findings (Schneider *et al*., 2006). On the other hand, the large effect found for systolic blood pressure in experiment 1 could not be replicated in experiment 2.

One might argue that the induction of expectancy has not worked properly. This was not the case: Participants believed our suggestions, as an analysis of expectancy measures and the debriefing showed. But these expectancies were rather uncorrelated with outcome. It might be that in a coffee drinking culture like Germany such effects are too small to become meaningful. On the other hand, it may well be that caffeine placebo effects unfold their effects primarily through carriers for which expectancies have been formed (i.e., coffee, tea, or cola).

A clear strength of our study distinguishing it from most studies published in the literature is that we used strictly double-blind methods as in our earlier work (Walach *et al*., 2001, 2002). This could explain why not even classical placebo effects were seen consistently. We engaged experimenters who were trained and

DOI: 10.1002/hup

knowledgeable in the experimental procedures but had no clue about our research hypotheses and the real substances. Whereas in all other studies, to our knowledge, at least one of the experimenters was privy to the information about study hypotheses, even if it remained undisclosed to the actual participants, in our experiment no such contact between participants and study designers ensued. None of the authors had any contact with any of the participants. Apart from cultural differences, this feature might explain why our placebo effects were rather low. It also demonstrates how crucial good blinding is in experiments and clinical trials in general.

When comparing the effects of the placebo groups of the double-blind arms in the two experiments, there was a medium sized difference for systolic blood pressure ($d = 0.58$), none for diastolic blood pressure ($d = 0.07$), small differences for the reaction time measure ($d = 0.25$; $d = 0.22$), and a small but negative effect size for alertness ($d = -0.31$). None of these differences was significant. We had conceived this as a test of a non-classical placebo effect, i.e., a difference in the magnitude of placebo effects due to the presence of a pharmaceutical substance in the second study. Although not formally significant, the effect sizes show that such a hypothesis cannot be excluded and might warrant further study. Due to the nature of the potential effect with positive and negative effect sizes of varying magnitude, a more sensitive analysis did not reveal any additional information.

A few words of caution are due: we did not conduct the experiment as a within-subject controlled study for principal reasons. Had we done so, we would have in effect conducted one experiment only, and caffeine would have been part of it right from the beginning, and for theoretical reasons we would have expected a correlation across experiments which would have not allowed us to make the comparisons we wanted to make. As we hypothesized that the replication, and the *introduction* of the substance in experiment 2, might have a differentiating effect, this method of control was not open to us.

Also, formally speaking, our results are not statistically significant. Although we saw strong pharmacological effects, the study was underpowered to demonstrate psychological placebo effects convincingly. By the same token, a more powerful replication study might be able to elucidate some of the questions left open by this first attempt.

One might argue that the introduction of caffeine in experiment 2 might have provided the experimenters with clues due to visual signs. Our debriefing did not reveal any hints, and the fact that differences between experiments are mainly visible in objective measures would speak against such an explanation.

In conclusion, our experiments show that placebo effects in an experimental analog to a randomized placebo-controlled trial are to some extent reproducible, but also quite variable. They are difficult to predict, and they might be modified by the presence of an active pharmacological substance.

CONFLICT OF INTERESTS

The authors have no conflicts of interest.

ACKNOWLEDGEMENTS

The study was supported by a grant from the Samueli Institute (www.siib.org), Alexandria VA. We are grateful to Katrin Hassenpflug and Lena Stopatschinskaja for conducting the experiments. We thank Rainer Trittler for providing the hydrochloride and caffeine solutions as well as Wolfgang Weinmann for performing the pharmacologic analyses.

REFERENCES

Amanzio M, Benedetti F. 1999. Neuropharmacological dissection of placebo analgesia expectation-activated opioid systems versus conditioning-activated specific subsystems. *J Neurosci* **19**: 484–494.

Amanzio M, Pollo A, Maggi G, Benedetti F. 2001. Response variability to analgesics: a role for non-specific activation of endogenous opioids. *Pain* **90**: 205–215.

Anderson C, Horne JA. 2008. Placebo response to caffeine improves reaction time performance in sleepy people. *Hum Psychopharmacol: Clin Exp* **23**: 333–336.

Andrews SE, Blumenthal TD, Flaten MA. 1998. Effects of caffeine and caffein-associated stimuli on the human startle eyeblink reflex. *Pharmacol Biochem Behav* **59**: 39–44.

Atmanspacher H, Römer H, Walach H. 2002. Weak quantum theory: complementarity and entanglement in physics and beyond. *Found Phys* **32**: 379–406.

Benedetti F. 1996. The opposite effects of the opiate antagonist naloxone and the cholecystokinin antagonist proglumide on placebo analgesia. *Pain* **64**: 535–543.

Benedetti F. 2002. How the doctor's words affect the patient's brain. *Eval Health Prof* **25**: 369–386.

Benedetti F, Amanzio M, Baldi S, *et al*. 1998. The specific effects of prior opioid exposure on placebo analgesia and placebo respiratory depression. *Pain* **75**: 313–319.

Chambers KP (ed.). 2009. *Caffeine and Health Research*. Nova Science Publishers: New York.

Cherkin DC, Sherman KJ, Avins AL, *et al*. 2009. A randomized trial comparing acupuncture, simulated acupuncture, and usual care for chronic low back pain. *Arch Intern Med* **169**: 858–866.

Cohen J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum: Hillsdale (Original erschienen 1977. Academic Press: New York).

Colloca L, Benedetti F. 2005. Placebos and painkillers: is mind as real as matter? *Nat Rev Neurosci* **6**: 545–552.

Colloca L, Fabrizio T, Recchia S, *et al*. 2008a. Learning potentiates neurophysiological and behavioral placebo analgesic responses. *Pain* **139**: 306–314.

Colloca L, Sigaudo M, Benedetti F. 2008b. The role of learning in nocebo and placebo effects. *Pain* **136**: 211–218.

de la Fuente-Fernández R, Schulzer M, Stoessl AJ. 2002. The placebo effect in neurological disorders. *Lancet Neurol* **1**: 85–91.

Enck P, Benedetti F, Schedlowski M. 2008. New insights into the placeo and nocebo responses. *Neuron* **59**: 195–206.

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 MT, Mulvihill LE, Vogel-Sprott M. 1994. The expected drug and its expected effect interact to determine placebo responses to alcohol and caffeine. *Psychopharmacology* **115**: 383–388.

Fillmore MT, Roach EL, Rice JT. 2002. Does caffeine counteract alcohol-induced impairment? The ironic effects of expectancy. *J. Stud. Alcohol* **63**: 745–754.

Flaten MA, Blumenthal TD. 1999. Caffeine-associated stimuli elicit conditioned response: an experimental model of the placebo effect. *Psychopharmacology* **145**: 105–112.

Haake M, Muller HH, Schade-Brittinger C, *et al*. 2007. German Acupuncture Trials (GERAC) for chronic low back pain: randomized, multicenter, blinded, parallel-group trial with 3 groups. *Arch Intern Med* **167**(17): 1892–1898.

Holm S. 1979. A simple sequentially rejective multiple test procedure. *Scand J Stat* **6**: 65–70.

Hróbjartsson A. 2002. What are the main methodological problems in the estimation of placebo effects? *J Clin Epidemiol* **55**: 430–435.

Hughes JR. 1992. Clinical importance of caffeine withdrawal. *N Engl J Med* **327**: 1160–1161.

Kaptchuk TJ, Kelley JM, Conboy LA, *et al*. 2008. Components of placebo effect: randomised controlled trial in patients with irritable bowel syndrome. *Br Med J* **336**: 999–1003.

Kirsch I, Sapirstein G. 1998. Listening to prozac but hearing placebo: a meta-analysis of antidepressant medication. *Prev Treat* **1**: 2a.

Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. 2008. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the food and drug administration. *PLoS Med* **5**(2): e45.

Kirsch I, Moore TJ, Scoboria A, Nicholls SS. 2002. The emperor's new drugs: an analysis of antidepressant medication data submitted to the U.S. Food and Drug Administration. *Prev Treat* **5**: Article 23.

Levine JD, Gordon NC, Fields HL. 1978. The mechanism of placebo analgesia. *Lancet* **II**: 654–657.

Lotshaw SC, Bradley JR, Brooks LR. 1996. Illustrating caffeine's pharmacological and expectancy effects utilizing a balanced placebo design. *J Drug Educ* **26**: 13–24.

Malani A, Houser D, Kevin M. 2008. Expectations mediate objective physiological placebo effects. *Adv Health Econ Health Serv Res: JAI* **20**: 311–327.

Mikalsen A, Bertelsen B, Flaten MA. 2001. Effects of caffeine, caffeine-associated stimuli, and caffeine-related information on physiological and psychological arousal. *Psychopharmacology* **157**: 373–380.

Moerman DE. 2002. *Meaning, Medicine, and the "Placebo Effect"*. Cambridge University Press: Cambridge.

Moerman DE, Jonas WB. 2002. Deconstructing the placebo effect and finding the meaning response. *Ann Intern Med* **136**: 471–476.

Pahlke F, König IR, Ziegler A. 2004. Randomization in Treatment Arms (RITA): Ein Randomisierungsprogramm für klinische Studien. *Informatik, Biometrie und Epidemiologie in Medizin und Biologie* **35**: 1–22.

Patel SM, Stason WB, Legedza A, *et al*. 2005. The placebo effect in irritable bowel syndrome trials: a meta-analysis. *Neurogastroenterol Motil* **17**: 332–340.

Pollo A, Amanzio M, Arlsanian A, Casadio C, Maggi G, Benedetti F. 2001. Response expectancies in placebo analgesia and their clinical relevance. *Pain* **93**: 77–84.

Pollo A, Torre E, Lopiano L, *et al*. 2002. Expectation modulates the response to subthalamic nucleus stimulation in Parkinsonian patients. *NeuroReport* **13**: 1383–1386.

Scharf H-P, Mansmann U, Streitberger K, *et al*. 2006. Acupuncture and knee osteoarthritis. *Ann Intern Med* **145**: 12–20.

Schneider R. 2009. Unspecific effects of caffeine consumption: when does the mind overrule the body? In *Caffeine and Health Research*, Chambers KP (ed.). Nova Science Publishers: New York; 143–160.

Schneider W, Eschman A, Zuccolotto A. 2002. *E-prime User's Guide*. Psychology Software Tools: Pittsburgh.

Schneider R, Grüner M, Heiland A, *et al*. 2006. Effects of expectation and caffeine on well-being, arousal, and reaction time. *Int J Behav Med* **13**: 330–339.

Steyer R, Schwenkmezger P, Notz P, Eid M. 1997. *Der Mehrdimensionale Befindlichkeitsfragebogen (MDBF)*. Hogrefe: Göttingen.

Stillfried Nv, Walach H. 2006. The whole and its parts: are complementarity and non-locality intrinsic to closed systems? *Int J Comput Anticipatory Syst* **17**: 137–146.

Svenningsson P, Nomikos GG, Ongini E, Fredholm BB. 1997. Antagonism of adenosine A2A receptors underlies the behavioural activating effect of caffeine and is associated with reduced experssion of messenger RNA for NGFI-A and NGFI-B in caudate-putamen and nucleus accumbens. *Neuroscience* **79**: 753–764.

Vase L, Riley JL, Price DD. 2002. A comparison of placebo effects in clinical analgesic trials versus studies of placebo analgesia. *Pain* **99**: 443–452.

Vickers AJ. 2003. Underpowering in randomized trials reporting a sample size calculation. *J Clin Epidemiol* **56**: 717–720.

Walach H. 2001. The efficacy paradox in randomized controlled trials of CAM and elsewhere: Beware of the placebo trap. *J Altern Complement Med* **7**: 213–218.

Walach H, Maidhof C. 1999. Is the placebo effect dependent on time? In *Expectancy, Experience, and Behavior*, Kirsch I (ed.). American Psychological Association: Washington, DC; 321–332.

Walach H, Sadaghiani C, Dehm C, Bierman DJ. 2005. The therapeutic effect of clinical trials: understanding placebo response rates in clinical trials—a secondary analysis. *BMC Med Res Methodol* **5**: 26.

Walach H, Schmidt S, Bihr Y-M, Wiesch S. 2001. The effects of a caffeine placebo and experimenter expectation on blood pressure, heart rate, well-being, and cognitive performance. *Eur Psychol* **6**: 15–25.

Walach H, Schmidt S, Dirhold T, Nosch S. 2002. The effects of a caffeine placebo and suggestion on blood pressure, heart rate, well-being and cognitive performance. *Int J Psychophysiol* **43**: 247–260.