

STUDY PROTOCOL

Open Access



A virtual experimenter to increase standardization for the investigation of placebo effects

Bjoern Horing^{1*}, Nathan D. Newsome², Paul Enck³, Sabarish V. Babu² and Eric R. Muth¹

Abstract

Background: Placebo effects are mediated by expectancy, which is highly influenced by psychosocial factors of a treatment context. These factors are difficult to standardize. Furthermore, dedicated placebo research often necessitates single-blind deceptive designs where biases are easily introduced. We propose a study protocol employing a virtual experimenter – a computer program designed to deliver treatment and instructions – for the purpose of standardization and reduction of biases when investigating placebo effects.

Methods: To evaluate the virtual experimenter's efficacy in inducing placebo effects via expectancy manipulation, we suggest a partially blinded, deceptive design with a baseline/retest pain protocol (hand immersions in hot water bath). Between immersions, participants will receive an (actually inert) medication. Instructions pertaining to the medication will be delivered by one of three metaphors: The virtual experimenter, a human experimenter, and an audio/text presentation (predictor "Metaphor"). The second predictor includes falsely informing participants that the medication is an effective pain killer, or correctly informing them that it is, in fact, inert (predictor "Instruction"). Analysis will be performed with hierarchical linear modelling, with a sample size of $N = 50$. Results from two pilot studies are presented that indicate the viability of the pain protocol ($N = 33$), and of the virtual experimenter software and placebo manipulation ($N = 48$).

Discussion: It will be challenging to establish full comparability between all metaphors used for instruction delivery, and to account for participant differences in acceptance of their virtual interaction partner. Once established, the presence of placebo effects would suggest that the virtual experimenter exhibits sufficient cues to be perceived as a social agent. He could consequently provide a convenient platform to investigate effects of experimenter behavior, or other experimenter characteristics, e.g., sex, age, race/ethnicity or professional status. More general applications are possible, for example in psychological research such as bias research, or virtual reality research. Potential applications also exist for standardizing clinical research by documenting and communicating instructions used in clinical trials.

Keywords: Virtual reality, Placebo mechanisms, Placebo analgesia, Bias, Blinding, Standardization

Background

To a large extent, placebo effects arise from the psychosocial setting in which a treatment is delivered, most notably the physician/patient or experimenter/participant interaction [1]. These effects can be substantial [2, 3]. However, it is difficult to systematically tease apart the contributions made by separate aspects of an interaction.

In this paper, we propose the use of a virtual experimenter as a platform to address these issues with minimized risk of bias.

We define the placebo effect (PE) as that part of an actually experienced symptom improvement mediated by psychological expectancy concerning a treatment, regardless of the treatment's actual efficacy [4]. This expectancy arises from multiple sources in the treatment's psychosocial context, such as verbal instruction [5], conditioning [6] or social learning [7].

* Correspondence: bhoring@clemsun.edu

¹Department of Psychology, Clemson University, Clemson, SC, USA
Full list of author information is available at the end of the article

PEs are most well known in the context of randomized clinical trials, where the placebo group receives an inert treatment to serve as control for the actual treatment [8]. However, the change happening in the placebo group is always a composite measure, including the actual placebo effect and numerous other factors [9, 10]. Some of these are a feature of the nosological subject matter, such as the natural course of disease and the fact that people tend to seek out treatment at peaks of symptom severity. These factors can lead to regression to the mean, which describes the fact that extreme values will tend to converge toward less extreme values at retest. Other factors include distinct psychological phenomena, like response biases, which cause study participants to modify their answering behavior after receiving a placebo treatment, without actually affecting their symptom experience [9, 10].

To investigate PEs per se, dedicated research involving additional control groups has been performed. Different from classical clinical trials where the placebo group is the control group, dedicated placebo research inquires into the conditions in which PEs occur, and the mechanisms involved [11]. A number of paradigms exist which employ active treatments [11, 12]. For example, in the “balanced placebo design” [13], the actual treatment (active vs. inert) is paired with the information with which it is administered (“active” vs. “inert”). Like double-blind clinical trials, this permits blinding. However, use of active treatment is often precluded due to ethical and regulatory obstacles. Especially when investigating healthy participants to identify basal mechanisms of the placebo response, this frequently leads to the employment of more basic designs. One of these designs is a single-blind deceptive placebo design. For example, in one of our own studies [14], two groups receiving the same inert treatment were compared, with one of them being (deceptively) told that the treatment is actually effective, and one being (truthfully) told that it is not. Under these circumstances, the above mentioned psychological confounds can be exacerbated. Not only are the investigated variables highly affected by psychological influences to begin with, but additionally, blinding of the experimenter is frequently impossible or impractical [10]. Consequently, while great progress has been made in identifying the neuropharmacological [15], neuronal [16] and even genetic [17] underpinnings of placebo responsiveness, some of this research stands on shaky foundations [10]. If the experimental interaction is not tightly controlled, it is difficult if not impossible to establish causation in placebo trials: The reported symptom change of a participant cannot merely be attributed to the (intentional) treatment difference, but possibly to subtle but systematic (unintentional) differences in the broader experimental context. These differences contribute to a mean difference of experimental group versus control

group. Some of these context factors have been explored, such as experimenter variables like a reassuring, positive attitude [18], empathy [1], affect-oriented communication [19], professional status [20] or sex [21]. However, these studies have been rarely replicated.

We propose that some of the limitations of research addressing psychological entities like symptom experiences could be overcome by the use of a virtual experimenter (VEx; Fig. 1). The VEx is a computer program designed to perform the function of randomized treatment allocation and treatment delivery. While preserving or even exceeding psychosocial characteristics of a conversation [22] (which are required for PEs to arise in the first place), it would perform social tasks with complete standardization – this is in contrast to pre-recorded messages, which do not induce the same extent of social presence as an embodied conversational agent.

From this framework, it is easy to manipulate aspects of the instruction or the appearance of the VEx without changing any of the other parameters, as would almost necessarily be the case if the task were performed by a human experimenter (HEX). While virtual reality has been used in some related applications before (e.g., as “virtual reality-based analgesic” distraction techniques [23] or interview partners [24] in the scope of psychological research), its use for the proposed purpose has not been investigated.

In the following, we present a comprehensive study protocol to test the VEx’ efficacy in eliciting PEs, to establish its viability as a substitute for treatment allocation and delivery by a HEX (section “Methods”). Due to

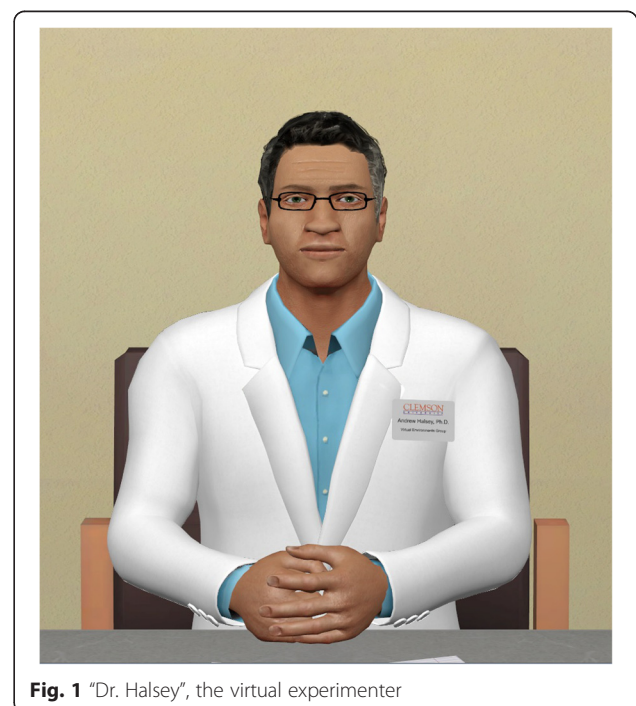


Fig. 1 “Dr. Halsey”, the virtual experimenter

the well-documented efficacy of placebo on pain (placebo analgesia), a pain modality will be employed as the method for symptom induction. The protocol combines several features (virtual reality software and pain protocol). These components have been separately investigated in pilot studies presented following the proposed protocol (section “Current status”). We outline the creation of the VEx under “Development of the stimulus material”, and its evaluation under “Pilot study 1”. To assess reproducibility and possible carry-over effects, we conducted a pilot study of the pain protocol, as well. Core findings are presented under “Pilot study 2”.

Methods

We propose the following methods to determine the VEx’ efficacy in inducing PEs. The interaction with the VEx will be compared to an HEx interaction, and an audio/text based interaction. All non-interventional interactions such as the pain testing will be performed by a blinded laboratory assistant.

Design

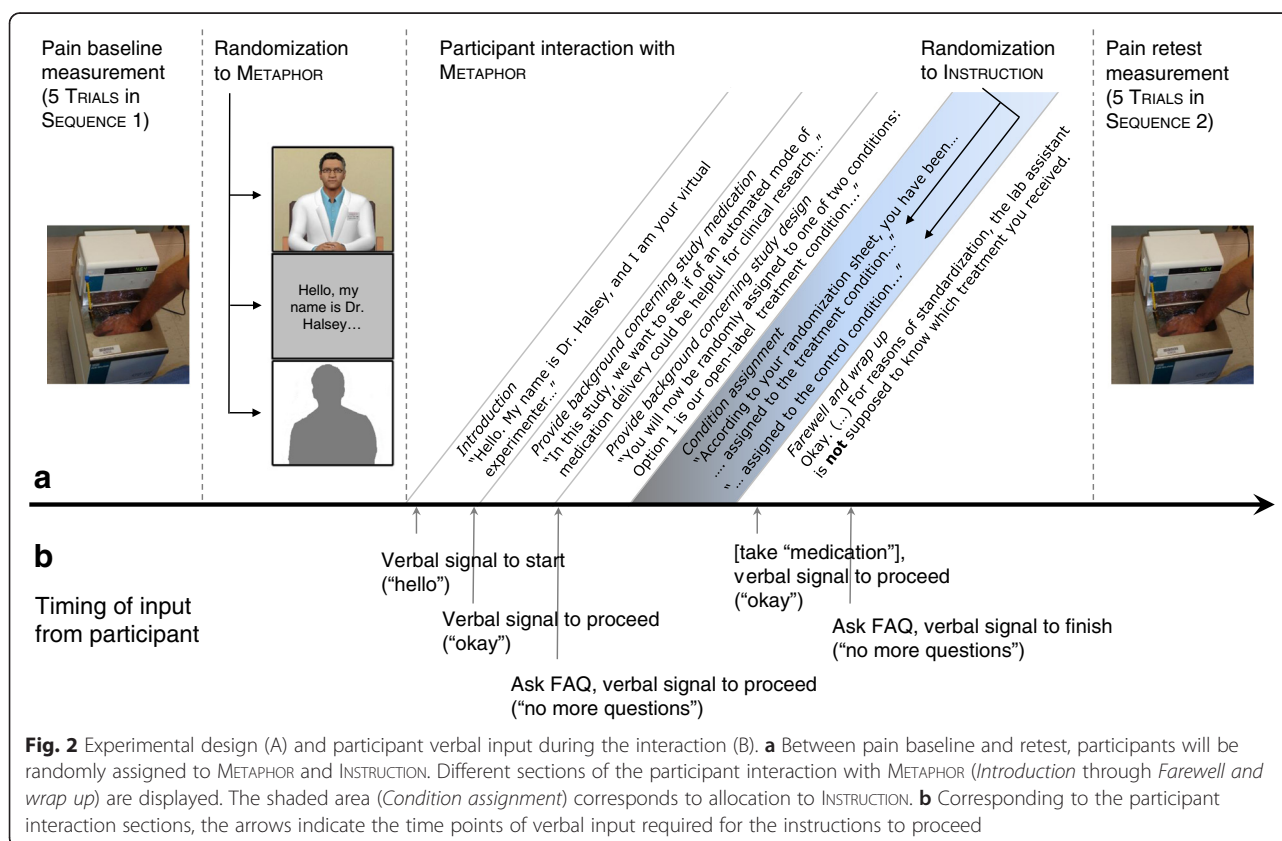
The study will follow a partially blinded, deceptive, repeated measures design. Participants will undergo a baseline pain measurement (five 60 s immersions of the hand in hot water; see “Pain protocol”), followed by treatment

with a supposed pain killer, and a retest pain measurement. In this design, every participant will only receive an oral placebo (no active medication), but groups will differ as to the instruction about the (actually inert) “medication’s” efficacy.

Within-persons and between-persons predictors include

- the mode of treatment/instruction presentation (between-persons predictor METAPHOR, 3 levels: HEx, VEx, text/audio)
- the instruction itself (between-persons predictor INSTRUCTION, 2 levels: “medication” (deceptive) vs. “inert” (truthful)); this manipulation is double-blind to both participant and laboratory assistant
- the pain measurement repetition, which constitutes nested within-persons predictors (immersion TIME nested in immersion TRIALS nested in immersion SEQUENCES). In the scope of this study outline, we will only address TRIALS and SEQUENCES pertaining to the expected results.

See Fig. 2a for a design and protocol outline. Within-persons and between-persons predictors correspond to level 1 and level 2 variables used in hierarchical linear modelling nomenclature (see “Analysis and sample size estimation”).



Study population

Participants will be recruited from Clemson University's student population, staff and faculty. Exclusion criteria will be applied (Table 1). We will balance the sample for sex/gender and will attempt mean matching for age, to yield roughly equal distributions in all combinations of the two between-persons predictors. Both sex/gender and age are possible confounds, as they may influence pain sensitivity [25, 26], placebo responsiveness [27] or the reception of the technology involved in the intervention [28].

Due to the deceptive nature of the protocol, full informed consent cannot be procured prior to participation; instead, a cover story will be presented to the participants. After the experiment, participants will receive full disclosure of the manipulation. Express permission will be asked of them to consider their data in the analysis. They will be given the choice to withdraw their data from analysis at any point prior to publication, without the need for justification.

Procedure

Participants will continually rate their pain during SEQUENCE 1 under supervision of a laboratory assistant. They will then be randomly assigned to one of the 3(METAPHOR) \times 2(INSTRUCTION) experimental groups. In a separate room, they will interact with the METAPHORS delivering the INSTRUCTIONS. For the VEx and audio/text condition, interactional components are displayed in Fig. 2b. Crucially, neither assignment will be revealed to the laboratory assistant.

During the placebo manipulation, the inert medication will be automatically dispensed using a self-constructed pill dispenser controlled by an Arduino platform (<https://www.arduino.cc/>). For VEx and audio/text condition, pill dispensation will be performed by the presentation software; in the HEx condition by pressing a button.

An overall inter-SEQUENCE interval of 35 min will be observed for cutaneous sensitivity to recover, and to provide a plausible cover story for the medication to take effect (in the "medication" instruction).

The pain change at retest, compared between experimental groups, will be the main outcome.

Table 1 Inclusion criteria for the proposed study

18 < Age < 64
No past or present brain or nerve conditions (e.g. fainting, stroke, neuropathy)
No past or present pain disorder (e.g. diagnosed migraine, chronic back pain)
No open wounds or skin conditions on the dominant hand (e.g. hangnail, graze)
Abstained from recreational drugs (alcohol, cannabis) for at least 24 h before the lab visit
Abstained from pain medication (and other medication feasibly affecting experimental parameters) for at least 24 h before the lab visit

Pain protocol

For symptom induction, we will employ a hot water bath (RTE-111, Neslab Instruments, Inc., Newington, NH), which is an established method for painful stimulation [29, 30]. Tonic heat pain has been successfully employed in the context of placebo studies with comparable stimulus strengths [31, 32]. Participants will be asked to immerse their dominant hand in the water kept at a constant temperature of 47 °C, which is painful for most people [33]. They will be informed that they can discontinue the experiment at any time. Two familiarization trials will be given where participants briefly put their hand in and out of the water. They will then be asked to immerse their hand for 1 min, take it out for 30 s, put it back in for 1 min etc., for a maximal number of 5 immersions (total immersion time 5 min). Immersion and removal will be prompted by automated verbal displays on screen, and accompanying audio signals.

Throughout the sequence, participants will continually rate their pain experience with a 0 to 100 digital visual analogue scale (VAS) displayed on a touch screen. They will be asked to leave their finger on the screen and move it once their pain perception changes.

After an interval of 35 min, another sequence of 5 immersions will be performed. After each sequence, they will also give a single retrospective rating of pain intensity and unpleasantness (VAS from 0 to 100) to establish comparability to conventional non-continuous measures.

Predictor METAPHOR: virtual experimenter, human experimenter and audio/text condition

Identical instructions will be delivered by the three METAPHORS. They will consist of a brief introduction of the agent delivering the instructions (VEx, HEx or disembodied voice), a cover story explaining the (actually fictional) background of the experiment, and the treatment allocation constituting the placebo manipulation. Participants will also interact with the VEx and audio/text condition in the scope of a frequently-asked questions (FAQ) module; the HEx will also provide the list of FAQs and will attempt to restrict the interaction to these. The questions revolve mostly around aspects of the experimental protocol, like they would in a naturalistic conversation with an HEx.

The VEx behavior includes looped idle animations, conversation-triggered gestures, and facial expressions. Experimental instructions were prerecorded by a voice actor and lip-synchronized with the virtual agent. The audio/text METAPHOR includes the same voice recording as the VEx METAPHOR (Fig. 3a), but displays text instructions alongside the recording, instead of the virtual agent (Fig. 3b). This condition is included to gauge the relevance of visual social cues beyond the verbal content of the instruction.

Additional details pertaining to the software employed in the experiment is provided below (see section "Current status").

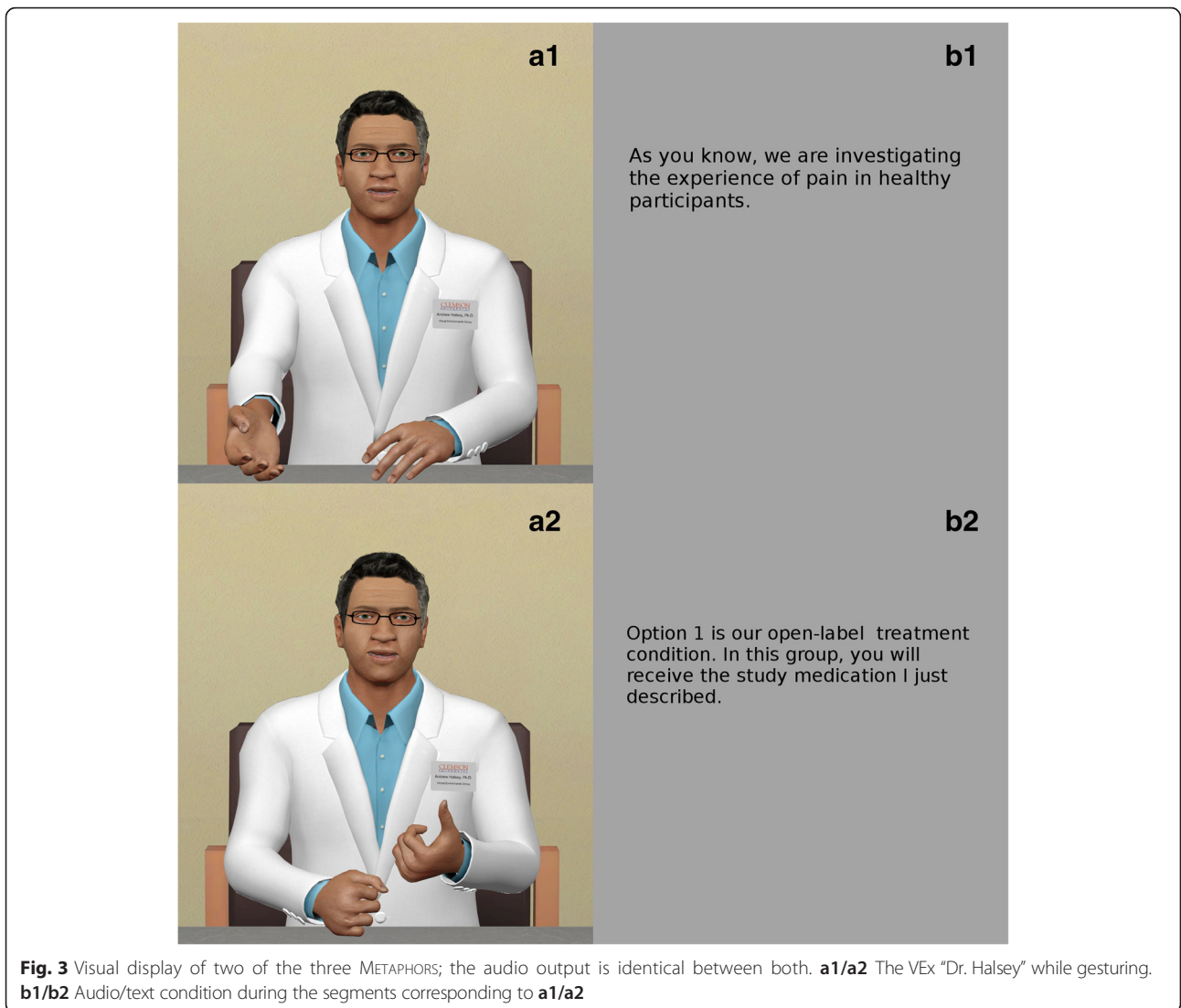


Fig. 3 Visual display of two of the three METAPHORS; the audio output is identical between both. **a1/a2** The VEx “Dr. Halsey” while gesturing. **b1/b2** Audio/text condition during the segments corresponding to **a1/a2**

Predictor INSTRUCTION: placebo manipulation versus control

The placebo manipulation is contained in the INSTRUCTION received by the participant around the dispensation of “medication”. It consists of the exchanging of a single sentence, stating either

“you have been assigned to the open-label treatment condition. This means that the pills you will receive are definitely the study medication” (in the “medication” INSTRUCTION)

or

“you have been assigned to the open-label control condition. This means that the pills you will take are placebo pills, that is, contain only an inactive substance” (in the “inert” INSTRUCTION).

To reinforce group allocation, two side effect questionnaires will be answered by the “medication” group only – one directly after placebo manipulation, the other at the end of the intermission. Furthermore, only “medication” group participants will be asked about their prior experience with the study drug, how effective it was in their experience, and how effective they assume it will be in reducing the pain at retest.

Secondary measures

A number of putative predictors of pain sensitivity and placebo responsiveness [34] will be assessed before and during the experiment. These variables will be used as adjunct within- and between-persons predictors (covariates) of changes in pain sensitivity.

Psychological measures include pain questionnaires as well as several other psychological trait and state

questionnaires (Table 2). Notably, participants will be asked about their desire for pain relief, and the expectation of pain relief after receiving the “medication” [5].

Physiological measures include heart rate as a proxy of general autonomic arousal [35] and high-frequency heart rate variability as a dedicated measure of parasympathetic withdrawal (cf. [36]). Furthermore, blood pressure [37] and skin temperature of the immersed hand [38] will be included as putative predictors of heat pain sensitivity. Heart rate and heart rate variability will be assessed with a Biolog device with Fetrodes technology (UFI, Morrow Bay, CA) during a 5 min physiological baseline, and during the measurements themselves (each sequence lasting 7.5 min). Blood pressure will be assessed with a calibrated GE Dinamap Pro100 (Medical Solutions, Minneapolis, MN) device. The last four of five measurements, spaced 1 min apart [39], will be averaged to derive systolic and diastolic blood pressure. Blood pressure will only be assessed during the physiological baseline, as cuff inflations could distract from the pain stimulation. Skin temperature will be assessed with an infrared thermometer before and after each SEQUENCE, as well as at multiple time points during the intermission.

Bias reduction and manipulation check

A manipulation check will be performed assessing any doubts or suspicions concerning the treatment allocation or the cover story. Preliminary data indicates that the cover story, including the group allocation, are considered

convincing by the participants: 18 of 19 people answered “no” to a post-experimental question asking “Did you at any point find the experimental instructions unconvincing?”.

Allocation to the experimental groups will happen in a room adjacent to the room where the pain measurement takes place, and without insight of the laboratory assistant into the allocation. He or she will therefore be fully blinded to the treatment allocation when performing the retest. Measures will be taken to ensure that allocation is not revealed by the participant: The participant will be asked by the VEx, HEx or audio/text display not to notify the laboratory assistant of their treatment allocation, and the laboratory assistant will document any deviations. Audiovisual recording of the participant/laboratory assistant-interaction will be considered if necessary.

All measurements (except blood pressure and skin temperature) and pain reports will be performed on a computer without communication with the laboratory assistant. Furthermore, the instructions for the continuous pain ratings contain a line to emphasize participant anonymity and non-involvement of the laboratory assistant in data analysis, “This next part is very important: Your ratings will be analyzed by a blinded experimenter only after the session is over. Focus as much as possible on the sensation in your immersed hand, and feel totally free to rate your pain precisely and truthfully. Try not to let my presence or any other thought affect the way you report the sensation, or affect the number of trials you complete.” Also, during the pain measurement, the

Table 2 Putative predictors of pain sensitivity and placebo responsiveness

Questionnaire/scale	Construct	Example item
Trait questionnaires, answered in online survey before coming to the lab		
Pain Related Self Statements inventory (PRSS)	Coping with pain, catastrophizing about pain	“If I stay calm and relax, things will be better.”
Pain Vigilance and Awareness Questionnaire (PVAQ)	Attentional focus on pain	“I am quick to notice changes in location or extent of pain.”
Pain Sensitivity Questionnaire (PSQ)	Sensitivity to various painful stimuli	“Imagine you trap your finger in a drawer.”
Beliefs about Medicines Questionnaire (BMQ)	Attitudes towards medical profession and pharmaceuticals	“Medicines do more harm than good.”
Beliefs in Expectation Biases	Beliefs to what extent perception is influenced by expectations	“In general, people are likely to experience the mood (good or bad) they expect to experience.”
International Personality Item Pool Big Five scales (IPIP B5)	Extraversion, agreeableness, conscientiousness, emotional stability, intelligence/imagination	“I feel comfortable around people.”
Internality, Powerful Others, and Chance scales (IPC)	Attribution of events to personal control, other individuals, chance	“Whether people act according to my wishes depends mainly on myself.”
Revised Life Orientation Test (LOT-R)	Positive/negative expectation of outcomes = dispositional optimism/pessimism	“I’m always optimistic about my future.”
State questionnaires, answered during the lab visit		
State Trait Anxiety Inventory (STAI)	Acute and general anxiety	“I get in a state of tension or turmoil as I think over my recent concerns and interests.”
Multidimensional Mood Questionnaire (MDMQ)	Good/bad mood, calmness/nervousness, wakefulness/tiredness	“Right now I feel [...] content.”

laboratory assistant will remove himself or herself from the immediate vicinity of the participant to sit down at a table in the participant's field of view, but facing in the same direction as the participant as to not give the participant the feeling of being observed (beyond an initial making sure that all instructions are complied with).

Some extent of selection bias is expected, as an aversive stimulation is employed. While a monetary compensation of \$20 is offered for the ca. 2 h experiment, pain sensitive people will be potentially underrepresented in the sample. Systematic deviations from the norm will be assessed by analyzing questionnaire data (specifically, reported pain sensitivity via the PSQ) and actual responding to the stimulus. Preliminary data does not suggest that this is a serious issue, with roughly normally distributed pain sensitivity (data not shown).

Analysis and sample size estimation

Data entry is almost fully automated via the touch-screen interface (for continuous pain ratings), a digital survey tool (for other pain ratings and psychological measurements), and on-line registration of physiological parameters. The exception are blood pressure and temperature data which are documented on a score sheet. Continuous pain ratings will be recorded at a sampling rate of ca. 40 Hz, but down-sampled to 1 s means for analysis.

Data will be analyzed with hierarchical linear modeling (HLM). Among other strengths, this approach enables us to consider both categorical and continuous predictor variables (e.g., covariates), and the inclusion of cases with missing values, allowing for the number of repeated measures to differ across persons [40]. HLM puts no constraints on the number of repeated measurements, which permits an analysis of differences in the time course of pain. HLM also allows for considering random effects (intercepts and slopes), suitable to the high interindividual variability of pain sensitivity [41].

The robustness of results will be assessed by outlier analyses of the regression models provided by HLM, including contrasting of predicted versus actual values, studentized residuals and Mahalanobis' distance. Outliers will be excluded both at the within-person level (individual measurement occasions) and the between-person level (exclusion of individuals with outliers). Robustness will be gauged by changes in the significance of the results.

HLM sample size estimation considers the sample size at level 1 (n ; in our case, the repeated measurements within-persons) and the sample size at level 2 (N ; in our case, the individual participants). Recommendations vary (e.g. [42, 43]); Hox [43] suggests larger sample sizes of $n = 20$ and $N = 50$ when investigating cross-level interactions. The wealth of data points provided by the second-by-second continuous pain ratings easily allow for meeting the $n = 20$ suggestion. However, while a total of 600 s of immersions

will be available, the effective sample size will be lower due to large intraclass correlations, or strong autocorrelation. Due to the high interindividual variation of pain sensitivity, we would nevertheless aim for including $N = 50$ individual participants to determine difference between the six groups. Estimating 15 % attrition, the projected sample size therefore amounts to $N = 58$ individuals.

Hypotheses and expected results

The PE would be constituted by a larger symptom improvement from SEQUENCE 1 to SEQUENCE 2 in the "medication" INSTRUCTION, than in the "control/placebo" INSTRUCTION ("medication"(S2-S1) > "control"(S2-S1)). This is a general test of the efficacy of the placebo paradigm, which is a crucial building block for assessing the effect of the different METAPHORS.

The design does not strictly permit the establishment of equivalence between METAPHORS in the sense of a non-inferiority trial, which is deemed too ambitious at this stage. For the proposed protocol, a non-significant difference between VEX and HEX would be considered as a first step to establish the viability of the VEX ($HEX = VEX$). On a descriptive level, we are expecting a slightly lower effect induced by the VEX, but lower variance than under the HEX, amounting to comparable effect sizes. Beyond that, it is hypothesized that the audio/text METAPHOR induces a smaller symptom improvement than the other conditions ($HEX = VEX >$ audio/text) due to absence of visual social cues.

Using a more fine grained analysis at the TRIAL level could yield additional information, such that groups differ in early or late TRIALS only, or that the slope (pain changes in consecutive TRIALS) is different between groups. For example, a significant difference between "medication" vs. "control" could come about by a shallower pain increase in consecutive TRIALS in SEQUENCE 2, despite identical initial sensitivity (see Fig. 4).

If we indeed do not find significant differences of PEs induced by the VEX versus an HEX, we will proceed to more complex protocols (see "Potential and outlook"). If we do find differences such that an HEX proves superior to the VEX, we will return to the design phase and address weaknesses in the VEX software in an iterative design-test protocol until we can reliably elicit PEs using the VEX that are of comparable size with those elicited by an HEX.

Current status

In the following, the development of the virtual experimenter is described in more detail. Furthermore, preliminary results from a pilot study pertaining to the VEX software are summarized (Pilot study 1). Finally, results from a pilot study pertaining to the pain protocol are summarized (Pilot study 2).

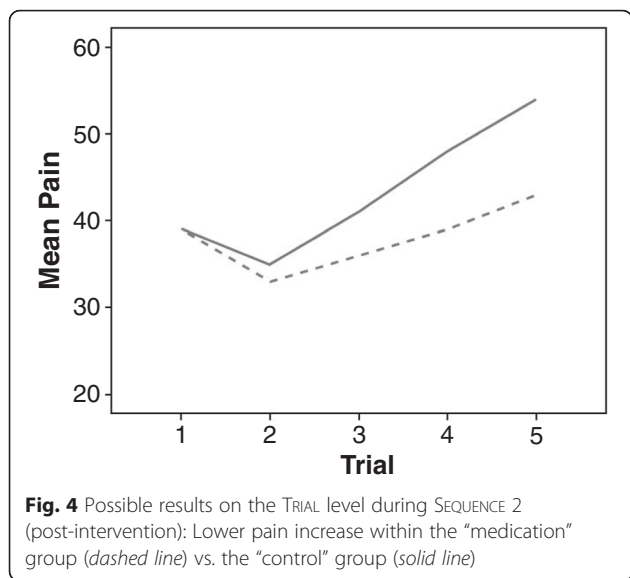


Fig. 4 Possible results on the TRIAL level during SEQUENCE 2 (post-intervention): Lower pain increase within the “medication” group (dashed line) vs. the “control” group (solid line)

Development of the stimulus material

The design of the VEx “Dr. Halsey” (cf. Fig. 1) was informed by a preliminary survey including eight different photographs of supposed study physicians. Photographs were rated by a non-random sample (N = 10) on various dimensions (e.g., appearing competent and empathetic [44], trustworthy, believable, and enthusiastic [1, 18]) (data not shown).

The VEx prototype was modeled using MakeHuman version 1.0.1 and Blender version 2.69 (Blender Foundation, Amsterdam, Netherlands). It is rendered in a Unity 5.1.1 framework (Unity Technologies, San Francisco, CA, USA) on a 60” screen display with 3D-capabilities. Lip-synchronization was performed using faceshift (faceshift, San Francisco, CA, USA). The software also includes an interaction module, which enables participants to ask the VEx a number of questions using a second

17” touch-screen display. Voice recognition is performed via the Microsoft Kinect (Microsoft, Redmond, WA) and processed using the Kinect Software Development Kit 1.8 (Microsoft, Redmond, WA) within the Unity framework rendering the VEx.

Pilot study 1: evaluation of the virtual experimenter

This study was aimed at informing the direction of further development of the VEx, by investigating how the VEx prototype is perceived by random participants. The study protocol outlined above will employ a version of the piloted prototype once all changes according to the findings of this study have been integrated.

In this pilot study, 48 participants (23 female, 25 male; age mean ± SD 20.2 ± 2.3) watched the VEx instruction without actually undergoing the pain protocol, or receiving the presumed medication. Participants judged the VEx on an Expectancy Induction Characteristics questionnaire assessing dimensions considered conducive to placebo responses (e.g. [1, 18, 19]). The Expectancy Induction Characteristics questionnaire was constructed in our lab and consists of 11 items (semantic differentials). The items are subsumed in three scales pertaining to how convincing a person is perceived (scale Message), how likeable (scale Likeable), and how compelling (scale Compelling). The scales exhibit decent reliability with Cronbach’s alpha of .82, .68 and .82, respectively.

Results are presented in Fig. 5. While the VEx was judged favorably on the Message and Likeability scales, it received comparatively bad rating on the Compelling scale. We believe this to be owed to a rather neutral demeanor (both in verbal and nonverbal content) displayed by the model, which has been intentional to establish a solid baseline of the VEx’s efficacy. However, we are currently in the process of improving ratings in this scale by implementing higher behavioral realism (e.g. breathing,

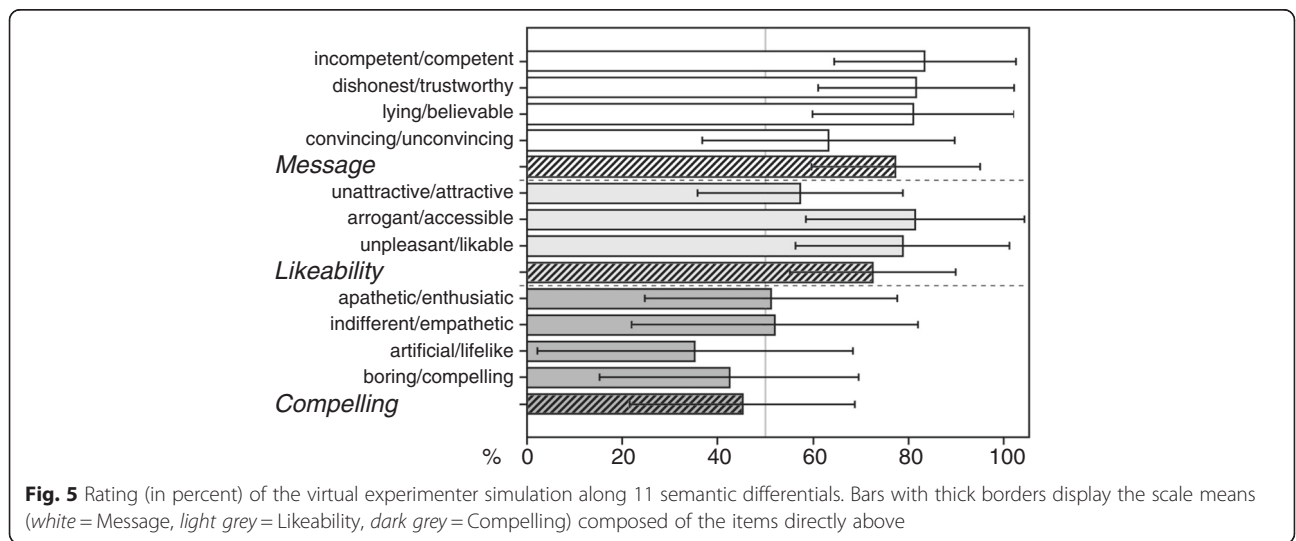


Fig. 5 Rating (in percent) of the virtual experimenter simulation along 11 semantic differentials. Bars with thick borders display the scale means (white = Message, light grey = Likeability, dark grey = Compelling) composed of the items directly above

pre-recorded eye blinks) and a more animated presentation (e.g. more inflections during speech, more pronounced facial expressions).

Pilot study 2: evaluation of the pain protocol

To characterize the pain protocol (see above) and estimate reproducibility across SEQUENCES (including possible carry-over effects like habituation and sensitization), we conducted a pilot study with only a waiting period between the two SEQUENCES. All means are provided with standard deviations; significance testing was performed with hierarchical linear model analysis, cf. “Analysis and sample size estimation” (a more in-depth discussion is beyond the scope of this article). In this study, we included 33 healthy participants (21 female, 12 male; age 20.1 ± 2.4). Continuous pain ratings almost invariably increased within SEQUENCES/between TRIALS. On average, participants showed substantial amounts of pain in both SEQUENCE 1 (VAS 46.4 ± 15.0 , range 23.5 to 83.6) and SEQUENCE 2 (VAS 40.8 ± 15.5 , range 18.8 to 78.8), with normal distributions on both occasions. These means constitute a slight but significant overall decrease in pain between the two SEQUENCES. Conversely, single retrospective ratings show no difference for intensity (VAS 53.8 ± 17.7 versus 52.6 ± 21.1) and unpleasantness (VAS 55.9 ± 17.1 versus 55.6 ± 21.3).

The continuous ratings exhibited good reproducibility (intraclass correlation coefficient(2,1) $\rho = .72$) between the two SEQUENCES, comparable to that of the retrospective ratings of intensity ($\rho = .77$) and unpleasantness ($\rho = .68$).

Discussion

Challenges

The two core challenges are the establishment of PEs in a tightly controlled protocol with only limited “dosing” of the placebo manipulation, and achieving a sufficient comparability of the VEx with the HEx interaction as to mimic the psychosocial effects necessary for PEs to arise. The placebo manipulation only contains a verbal expectancy manipulation comparable to those used in other placebo experiments (e.g. [5]). However, we chose to limit emphasis and enthusiasm on the side of the experimenter as much as possible, as to introduce as few as possible confounding factors at this point. This may mean that the PE will not be pronounced.

With the VEx, the greatest challenge is to create a sufficient rapport between him and the participant to be perceived as a social situation. The Ethopoeia concept [45] posits a substrate neutrality of human versus virtual agent behavior: as long as social cues are displayed, they will lead to “mindless responses” [45], p. 83). However, these cues go beyond the behavioral realism of the display (e.g. facial expressions, gestures, blinking, breathing) – which is necessary but not sufficient [46] – and extent to interactional components to attribute agency to the

virtual agent, such as turn-taking, visual attention to interlocutors, and behaviors reducing inter-personal distance. While technology like the “rapport agent” exists [47, 48], which has been tailored to naturally engage humans in conversation, our VEx software does not yet exhibit the same capabilities. At this stage, the VEx includes an FAQ module, engaging participants to ask questions with prerecorded answers; furthermore, verbal input will be required from the participants at certain points of the protocol to continue (cf. Fig. 3b). However, in future iterations, actually artificial intelligence-driven conversation is planned [49], engaging participants in personal questions e.g. about demographic variables in a small talk/rapport module.

As to the appearance of the VEx, a balance will be struck between realism and acceptance – specifically, the design has to avoid the so-called “uncanny valley” [50] describing an aversive perception of almost-but-not-quite realistic artificial agents by human spectators.

Limitations

The rationale of using the VEx in the context of experimental studies makes the assumptions that the treatment delivery and the accompanying instructions are, in fact, the most important aspects for the generation of PRs. While instructions alone have been shown to effect PEs [5, 51, 52], it has been demonstrated that conditioning paradigms may have better results [6]. However, a conditioning approach was considered too logistically demanding at this stage.

It is possible that the control group itself will not fully believe in the veracity of them only receiving inert medication. However, this would lead to a conservative estimate of placebo effectiveness and is therefore no hazard to the internal validity of the study. We are also considering the option of using a no-treatment control (i.e., no placebo delivery at all) instead of a “control/placebo” instruction.

While evidence exists that female doctors are estimated to have better explanation skills and technical skills and have a higher acceptance [53], the decision to only include male models in the creation of the VEx was made for opportunistic reasons (having a male voice actor available).

On a more general note, while the main intention behind the creation of the VEx is the assessment and abolishment of biases in psychological and medical research, there is a possibility that the use of a virtual agent would induce biases of another sort, such as involving computer (il)literacy or opinions towards technological progress [54]. For example, it is feasible that older participants would have a different response to the VEx than younger participants, who will have been exposed to virtual reality technology to a higher extent. On the plus side, there is

evidence that participants tend to disclose more personal information to a virtual agent than a human experimenter [55, 56], which may be of relevance with increased interactional capabilities of the VEx. Nevertheless, generalizability will be attempted by using a sample beyond student participants. We are further addressing these issues by post-experiment questionnaires and demographical data about the experience with, and daily use of computers and 3D technology.

Furthermore, the novelty of the simulation may serve to accentuate or alleviate effects which would otherwise be (in)effective. The experience of novelty will be assessed and investigated as a potential moderator of treatment efficacy.

Potential and outlook

The broader objective of investigating the VEx is to estimate its comparability to a real human. Using a placebo paradigm is considered an optimal testbed of this comparability due to its dependence on psychosocial variables. Regardless of outcome pertaining to the efficacy of the VEx, results from this study will be disseminated to a wider public, since they will contain valuable information on the specific mechanisms involved in the generation of PEs. Success of the study therefore does not hinge on the efficacy of the VEx, only on replicating results from previous placebo manipulations by the HEx.

Should it be established that the VEx can, in fact, serve as a substitute to an HEx, the possible applications in medical and general psychological research are many. Furthermore, it should be kept in mind that while a VEx would prove to be of great value for placebo research, applicability goes far beyond this narrow topic and extends into most forms of psychological experiments where standardization is of relevance. These applications fall into three interrelated categories – standardization, experimental modification, and communication. To these ends, we are making efforts to develop the software such that it can be easily adopted by other researchers, for whom it would eventually be made available.

Standardization of experimenter characteristics would facilitate the investigation of psychological entities feasibly affected by psychosocial interactions. As no repeated acting/performance would be required, a VEx could serve as a “perfect confederate” [57], p. 6) in social psychological studies, or studies involving social learning. A VEx would get rid of experimenter-side biases due to unblinding, especially if being a stakeholder in an experiment’s success or other factors would incentivize an HEx to guess treatment allocations or behave differently according to which group one is assigned to, introducing experimenter demand characteristics [58].

For placebo research, the putative role of experimenter variables has been elaborated above. Yet, these factors have been underreported [59], introducing confounders

and thus exacerbating the problems created under non-blind conditions. One reason for this is the difficulty of assessing or standardizing these factors. Not only could a VEx serve as a perfectly standardized platform if these factors are sought to be held constant: Instead, it would provide a convenient means to modify single aspects of the experimenter characteristics while holding all others equal – such as gender, age or professional role of the experimenter, or informational content such as the phrasing of instructions. This would not only inform the respective research (e.g. about placebo mechanisms), but also the creation of virtual agents used as interfaces in customer relations and medical devices.

Furthermore, with a VEx interface, participant behavior could be monitored much easier, for example by logging which questions are asked by the participant. While it will strongly depend on the concrete experimental context if a VEx can exhaustively serve the informational needs of a participant, pilot studies could inform the formulation of frequently asked questions to be covered by the VEx.

Lastly, the standardization aspect not only applies to single experiments, but to collaborative, multi-center research efforts as well. Since the communication effectively only entails distribution of software, it could serve as an easy means to establish instruction equivalence. This is particularly relevant considering that insufficient reliability poses serious economic and scientific problems for randomized controlled trials [60, 61], with even double-blind conditions being no safeguard against yielding biased results [62]. It is conceivable that outsourcing suitable study tasks (like consent, instructions, medication, questionnaire administration) would free up resources which could be spent on other activities, reducing a trial’s logistical burden. The VEx could be used as best practice example or in training videos. It could contribute to advances in telemedicine or similar situations where a virtual human interface has been shown to increase usability, e.g., with elderly people [63]. With the preservation of social cues, the possibility of instruction delivery via different telecommunication devices could prove invaluable for ambulatory treatment where improvement in compliance and treatment efficacy are sought.

Abbreviations

FAQ, frequently asked questions; HEx, human experimenter; HLM, hierarchical linear modelling; PE, placebo effect; VAS, visual analogue scale; VEx, virtual experimenter

Acknowledgements

Not applicable.

Funding

This research has been supported by an intramural grant by Clemson University’s Transformative Initiative for Generating Extramural Research (TIGER). It was further supported by a Feodor-Lynen Research Fellowship by the Alexander von Humboldt-Foundation, Germany, awarded to BH. The funding bodies were not involved in study design, or collection, analysis and interpretation of data, or the writing of the manuscript.

Availability of data and materials

The pilot datasets analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

BH conceived of the study, participated in the study design, coordinated data collection (VEx and pain protocol pilot study) and drafted the manuscript. NN developed the study software and conducted data collection for its evaluation (VEx pilot study). SB contributed to study conception and participated in the design of the study software. PE participated in the design of the study software and helped to draft the manuscript. EM contributed to study conception, participated in the study design, and helped to draft the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

The proposed experiment as well as all pilot studies reported here were reviewed by Clemson University's institutional review board, and considered to conform to ethical standards (vote IRB2014-254). For both the proposed experiment and the pilot studies, participants will sign or signed full informed consent prior to participation.

An ethical concern of placebo mechanism research is that most paradigms require some form of deception. At first glance, deception is at odds with tenets of medical ethics, which emphasize the empowerment of the patient to make informed decisions about his or her treatment [64] – the same applies, albeit to a lesser extent, to the involvement of healthy participants [65]. However, the use of placebos is clearly indicated in situations where truthful information would make valid results impossible – hence, the Declaration of Helsinki accepts deception “where for compelling and scientifically sound methodological reasons the use of placebo is necessary to determine the efficacy or safety of an intervention and the patients who receive placebo or no treatment will not be subject to any risk of serious or irreversible harm...” [66]. Identical reasons are provided by the American Psychological Association [65]. While some research indicates that deception authorized by the participants (“authorized deception”) [67, 68] or even placebo *without* deception (“open-label”) [69] are viable under certain circumstances, the novelty of the paradigm employed suggests to rely on a more classical, deceptive approach concerning the nature of the conditioning procedure, with full disclosure afterwards. Incidentally, even in clinical settings, the use of deception is less of a concern to the actual participants [70], than a matter of principle for medical ethics. We emphasize that the proposed research is in line with common practice of institutional review boards, and that the acceptance of this form of deception has been demonstrated in pilot studies.

Should the VEx software prove viable for the current purpose, and a customizable version be made available to a broader (presumably academic) audience, it will be accompanied by appropriate caveats concerning its generalizability, and that it cannot possibly be a one-size-fit-all substitute for “the human in the loop” for the sake of placebo controls. Ethical approval will have to be obtained for each specific instance of its employment. These caveats seem to be necessary, as exemplified by a superfluity (and consequently uncritical use) of mHealth applications [71], which have only recently been subjected to regulations in the form of guidelines by regulatory bodies in Europe and the USA [72, 73].

Author details

¹Department of Psychology, Clemson University, Clemson, SC, USA.

²Human-Centered Computing Division, School of Computing, Clemson University, Clemson, SC, USA. ³Department of Psychosomatic Medicine and Psychotherapy, University Hospital Tübingen, Tübingen, Germany.

Received: 14 October 2015 Accepted: 7 July 2016

Published online: 18 July 2016

References

1. Kelley JM, Lembo AJ, Ablon JS, Villanueva JJ, Conboy LA, Levy R, et al. Patient and practitioner influences on the placebo effect in irritable bowel syndrome. *Psychosom Med.* 2009;71(7):789–97.
2. Schedlowski M, Enck P, Rief W, Bingel U. Neuro-bio-behavioral mechanisms of placebo and nocebo responses: implications for clinical trials and clinical practice. *Pharmacol Rev.* 2015;1328:1–34.
3. Vase L, Petersen GL, Riley JL, Price DD. Factors contributing to large analgesic effects in placebo mechanism studies conducted between 2002 and 2007. *Pain.* 2009;145(1–2):36–44.
4. Finniss DG, Kaptchuk TJ, Miller FG, Benedetti F. Biological, clinical, and ethical advances of placebo effects. *Lancet.* 2010;375(9715):686–95.
5. Vase L, Robinson ME, Verne GN, Price DD. The contributions of suggestion, desire, and expectation to placebo effects in irritable bowel syndrome patients: an empirical investigation. *Pain.* 2003;105(1–2):17–25.
6. Colloca L, Petrovic P, Wager TD, Ingvar M, Benedetti F. How the number of learning trials affects placebo and nocebo response. *Pain.* 2011;151(2):430–9.
7. Hunter T, Siess F, Colloca L. Socially induced placebo analgesia: a comparison of a pre-recorded versus live face-to-face observation. *Eur J Pain.* 2014;18(7):914–22.
8. Kaptchuk TJ. The double-blind, randomized, placebo-controlled trial: gold standard or golden calf? *J Clin Epidemiol.* 2001;54(6):541–9.
9. Kienle GS, Kiene H. The powerful placebo effect: fact or fiction? *J Clin Epidemiol.* 1997;50(12):1311–8.
10. Hróbjartsson A, Kaptchuk TJ, Miller FG. Placebo effect studies are susceptible to response bias and to other types of biases. *J Clin Epidemiol.* 2011;64(11):1223–9.
11. Enck P, Bingel U, Schedlowski M, Rief W. The placebo response in medicine: minimize, maximize or personalize? *Nat Rev Drug Discov.* 2013;12(3):191–204.
12. Enck P, Klosterhalfen S, Weimer K, Horing B, Zipfel S. The placebo response in clinical trials: more questions than answers. *Philos Trans R Soc Lond B Biol Sci.* 2011;366(1572):1889–95.
13. Rohsenow DJ, Marlatt GA. The balanced placebo design: methodological considerations. *Addict Behav.* 1981;6(2):107–22.
14. Horing B, Weimer K, Schrade D, Muth ER, Scisco JL, Enck P, et al. Reduction of motion sickness with an enhanced placebo instruction: an experimental study with healthy participants. *Psychosom Med.* 2013; 75(1534–7796):497–504.
15. Benedetti F, Amanzio M, Rosato R, Blanchard C. Nonopioid placebo analgesia is mediated by CB1 cannabinoid receptors. *Nat Med.* 2011;17(10):1228–30.
16. Bingel U, Tracey I, Wiech K. Neuroimaging as a tool to investigate how cognitive factors influence analgesic drug outcomes. *Neurosci Lett.* 2012; 520(2):149–55.
17. Hall KT, Loscalzo J, Kaptchuk TJ. Genetics and the placebo effect: the placeboome. *Trends Mol Med.* 2015;21(5):285–94.
18. di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijnen J. Influence of context effects on health outcomes: a systematic review. *Lancet.* 2001;357(9258):757–62.
19. Verheul W, Sanders A, Bensing J. The effects of physicians' affect-oriented communication style and raising expectations on analogue patients' anxiety, affect and expectancies. *Patient Educ Couns.* 2010;80(3):300–6.
20. Kállai I, Barke A, Voss U. The effects of experimenter characteristics on pain reports in women and men. *Pain.* 2004;112(1–2):142–7.
21. Aslaksen PM, Myrbakk IN, Høifødt RS, Flaten MA. The effect of experimenter gender on autonomic and subjective responses to pain stimuli. *Pain.* 2007; 129(3):260–8.
22. Gratch J, Wang N, Okhmatovskaia A, Lamothe F, Morales M, van der Werf R, et al. Can virtual humans be more engaging than real ones? In: Jacko J, editor. *Human-Computer Interaction, HCI Intelligent Multimodal Interaction Environments*, Lecture Notes in Computer Science 4552. Berlin: Springer; 2007. p. 286–297.
23. Triberti S, Repetto C, Riva G. Psychological factors influencing the effectiveness of virtual reality-based analgesia: a systematic review. *Cyberpsychol Behav Soc Netw.* 2014;17(6):335–45.
24. Hasler BS, Tuchman P, Friedman D. Virtual research assistants: replacing human interviewers by automated avatars in virtual worlds. *Comput Human Behav.* 2013;29(4):1608–16.
25. Racine M, Tousignant-Laflamme Y, Kloda LA, Dion D, Dupuis G, Choinire M. A systematic literature review of 10 years of research on sex/gender and experimental pain perception - Part 1: are there really differences between women and men? *Pain.* 2012;153(3):602–18.
26. Ledowski T, Stein J, Albus S, MacDonald B. The influence of age and sex on the relationship between heart rate variability, haemodynamic variables and

- subjective measures of acute post-operative pain. *Eur J Anaesthesiol*. 2011; 28(6):433–7.
27. Bjorkedal E, Flaten MA. Expectations of increased and decreased pain explain the effect of conditioned pain modulation in females. *J Pain Res*. 2012;5:289–300.
 28. Felnhöfer A, Kothgassner OD, Beutl L, Hlavacs H, Kryspin-Exner I. Is virtual reality made for men only? Exploring gender differences in the sense of presence. In: Annual conference of the international society on presence research. 2012.
 29. Horn C, Schaller J, Lautenbacher S. Investigating the affective component of pain: no startle modulation by tonic heat pain in startle responsive individuals. *Int J Psychophysiol*. 2012;84(3):254–9.
 30. Rainville P, Duncan GH, Price DD, Carrier B, Bushnell MC. Pain affect encoded in human anterior cingulate but not somatosensory cortex. *Science*. 1997;277(5328):968–71.
 31. Matre D, Casey KL, Knardahl S. Placebo-induced changes in spinal cord pain processing. *J Neurosci*. 2006;26(2):559–63.
 32. Lyby PS, Aslaksen PM, Flaten MA. Is fear of pain related to placebo analgesia? *J Psychosom Res*. 2010;68(4):369–77.
 33. Magerl W, Krumova EK, Baron R, Tölle T, Treede RD, Maier C. Reference data for quantitative sensory testing (QST): refined stratification for age and a novel method for statistical comparison of group data. *Pain*. 2010;151(3):598–605.
 34. Horing B, Weimer K, Muth ER, Enck P. Prediction of placebo responses: a systematic review of the literature. *Front Psychol*. 2014;5(October):1–10.
 35. Loggia ML, Juneau M, Bushnell MC. Autonomic responses to heat pain: heart rate, skin conductance, and their relation to verbal ratings and stimulus intensity. *Pain*. 2011;152(3):592–8.
 36. Streff A, Kuehl LK, Michaux G, Anton F. Differential physiological effects during tonic painful hand immersion tests using hot and ice water. *Eur J Pain*. 2010;14(3):266–72.
 37. Saccò M, Meschi M, Regolisti G, Detrenis S, Bianchi L, Bertorelli M, et al. The relationship between blood pressure and pain. *J Clin Hypertens*. 2013;15(8):600–5.
 38. Wu G, Campbell JN, Meyer RA. Effects of baseline skin temperature on pain ratings to suprathreshold temperature-controlled stimuli. *Pain*. 2001;90(1–2):151–6.
 39. Myers MG, Valdivieso M, Kiss A. Optimum frequency of office blood pressure measurement using an automated sphygmomanometer. *Blood Press Monit*. 2008;13(6):333–8.
 40. Raudenbush SW, Bryk AS. Hierarchical linear models: applications and data analysis. Thousand Oaks: Sage; 2002.
 41. Nielsen CS, Price DD, Vassend O, Stubhaug A, Harris JR. Characterizing individual differences in heat-pain sensitivity. *Pain*. 2005;119(1–3):65–74.
 42. Tabachnick BG, Fidell LS. Using multivariate statistics. Harlow: Pearson; 2014.
 43. Hox JJ. Multilevel analysis: techniques and applications. 2nd ed. New York: Routledge; 2010.
 44. Fiske ST, Cuddy AJC, Glick P. Universal dimensions of social cognition: warmth and competence. *Trends Cogn Sci*. 2007;11(2):77–83.
 45. Nass C, Moon Y. Machines and mindlessness: social responses to computers. *J Soc Issues*. 2000;56(1):81–103.
 46. Blascovich J. A theoretical model of social influence for increasing the utility of collaborative virtual environments. In: Proceedings of the 4th international conference on collaborative virtual environments. New York: ACM Press; 2002. p. 25–30.
 47. Gratch J, Okhmatovskaia A, Lamothe F, Marsella S, Morales M, van der Werf RJ, et al. Virtual rapport. In: Gratch J, Young M, Aylett R, Ballin D, Olivier P, editors. Intelligent virtual agents. Berlin, Heidelberg: Springer; 2006. p. 14–27.
 48. Huang L, Morency L-P, Gratch J. Virtual rapport 2.0. In: Vilhjálmsón HH, Kopp S, Marsella S, Thórisson KR, editors. Intelligent virtual agents. Heidelberg: Springer; 2011. p. 68–79.
 49. Babu SV, Schmutz S, Barnes T, Hodges LF, Babu SV, Schmutz S, Barnes T, Hodges LF. "What would you like to talk about?" an evaluation of social conversations with a virtual receptionist. In: Gratch J, Young M, Aylett R, Ballin D, Olivier P, editors. Intelligent virtual agents. Heidelberg: Springer; 2006. p. 169–80.
 50. Mori M, MacDorman K, Kageki N. The uncanny valley [from the field]. *IEEE Robot Autom Mag*. 2012;19(2):98–100.
 51. Price DD, Milling LS, Kirsch I, Duff A, Montgomery GH, Nicholls SS. An analysis of factors that contribute to the magnitude of placebo analgesia in an experimental paradigm. *Pain*. 1999;83(2):147–56.
 52. Elsenbruch S, Kotsis V, Benson S, Rosenberger C, Reidick D, Schedlowski M, et al. Neural mechanisms mediating the effects of expectation in visceral placebo analgesia: an fMRI study in healthy placebo responders and nonresponders. *Pain*. 2012;153(2):382–90.
 53. Shah R, Ogdan J. "What's in a face?" the role of doctor ethnicity, age and gender in the formation of patients' judgements: an experimental study. *Patient Educ Couns*. 2006;60(2):136–41.
 54. Bagozzi RP, Davis FD, Warshaw PR. Development and test of a theory of technological learning and usage. *Hum Relations*. 1992;45(7):659–86.
 55. Weisband S, Kiesler S. Self disclosure on computer forms. In: Proceedings of the SIGCHI conference on human factors in computing systems. New York: ACM Press; 1996. p. 3–10.
 56. Joinson AN. Self-disclosure in computer-mediated communication: the role of self-awareness and visual anonymity. *Eur J Soc Psychol*. 2001;31(2):177–92.
 57. Gratch J, Marsella S. Social emotions in nature and artifact. New York: Oxford University Press; 2013. p. 336.
 58. Orne MT. On the social psychology of the psychological experiment: with particular reference to demand characteristics and their implications. *Am Psychol*. 1962;17(11):776–83.
 59. Hughes J, Gabbay M, Funnell E, Dowrick C. Exploratory review of placebo characteristics reported in randomised placebo controlled antidepressant drug trials. *Pharmacopsychiatry*. 2012;45(1):20–7.
 60. Collier R. Rapidly rising clinical trial costs worry researchers. *CMAJ*. 2009;180(3):277–8.
 61. de Carvalho ECA, Jayanti MK, Batilana AP, Kozan AMO, Rodrigues MJ, Shah J, et al. Standardizing clinical trials workflow representation in UML for international site comparison. *PLoS One*. 2010;5(11):e13893.
 62. Hróbjartsson A, Forfang E, Haahr MT, Als-Nielsen B, Brorson S. Blinded trials taken to the test: an analysis of randomized clinical trials that report tests for the success of blinding. *Int J Epidemiol*. 2007;36(3):654–63.
 63. Kenny P, Parsons T, Gratch J, Rizzo A. Virtual humans for assisted health care. In: Proceedings of the 1st ACM International Conference on Pervasive Technologies Related to Assistive Environments. New York: ACM Press; 2008. p. 1.
 64. Beauchamp TL, Childress JF. Principles of biomedical ethics. 7th ed. New York: Oxford University Press; 2013.
 65. American Psychological Association. Ethical principles of psychologists and code of conduct [Internet]. 2010. Available from: <http://apa.org/ethics/code/index.aspx>. Accessed 1 July 2016.
 66. World Medical Association. WMA Declaration of Helsinki - Ethical principles for medical research involving human subjects [Internet]. 2008. Available from: <http://www.wma.net/en/30publications/10policies/b3/>. Accessed 1 July 2016.
 67. Miller FG, Wendler D, Swartzman LC. Deception in research on the placebo effect. *PLoS Med*. 2005;2(9):e262.
 68. Martin AL, Katz J. Inclusion of authorized deception in the informed consent process does not affect the magnitude of the placebo effect for experimentally induced pain. *Pain*. 2010;149(2):208–15.
 69. Schafer SM, Colloca L, Wager TD. Conditioned placebo analgesia persists when subjects know they are receiving a placebo. *J Pain Elsevier Ltd*. 2015; 16(5):412–20.
 70. Kisaalita N, Staud R, Hurlley R, Robinson M. Placebo use in pain management: the role of medical context, treatment efficacy, and deception in determining placebo acceptability. *Pain*. 2014;155(12):2638–45.
 71. Cortez NG, Cohen IG, Kesselheim AS. FDA regulation of mobile health technologies. *N Engl J Med*. 2014;371(4):372–9.
 72. European Commission. Green paper on mobile health ("mHealth") (COM(2014) 219 final) [Internet]. 2014. Available from: <https://ec.europa.eu/digital-single-market/en/news/green-paper-mobile-health-mhealth>. Accessed 1 July 2016.
 73. US Food and Drug Administration (FDA). Mobile Medical Applications [Internet]. 2015. Available from: <http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM263366.pdf>. Accessed 1 July 2016.