

Highlights

The Placebo Effect of Human Augmentation: Anticipating Cognitive Augmentation Increases Risk-Taking Behavior

- Expectations of improvement for human augmentation technology affect evaluation even if the technology is nonfunctional.
- Expectations of improvement for human augmentation technology increase risk-taking behavior.
- Participants processed information differently for the non-functional augmentation when encountering a loss card.

The Placebo Effect of Human Augmentation: Anticipating Cognitive Augmentation Increases Risk-Taking Behavior

Steeven Villa^{a,*1} (Researcher), Thomas Kosch^b, Felix Grelka^a, Albrecht Schmidt^a and Robin Welsh^c

^aLMU Munich, Frauenlobstrasse 7a, Munich, 80337, Germany

^bHU Berlin, , Berlin, 80337, Germany

^cAalto University, , Espoo, 80337, Finland

ARTICLE INFO

Keywords:
Placebo Effect
Human Augmentation
User Expectations
Electroencephalography
User Studies

ABSTRACT

Human Augmentation Technologies improve human capabilities using technology. In this study, we investigate the placebo effect of Augmentation Technologies. Thirty naïve participants were told to be augmented with a cognitive augmentation technology or no augmentation system while conducting a Columbia Card Task. In this risk-taking measure, participants flip win and loss cards. The sham augmentation system consisted of a brain-computer interface allegedly coordinated to play non-audible sounds that increase cognitive functions. However, no sounds were played throughout all conditions. We show a placebo effect in human augmentation, where a sustained belief of improvement after using the sham system remains and an increase in risk-taking conditional on heightened expectancy using Bayesian statistical modeling. Furthermore, we identify differences in event-related potentials in the electroencephalogram that occur during the sham condition when flipping loss cards. Finally, we integrate our findings into theories of human augmentation and discuss implications for the future assessment of augmentation technologies.

1. Introduction

Human Augmentation Technologies (ATs) are ubiquitous near-body technologies that enable users to improve their cognitive, sensory, or physical abilities. ATs enhance human capabilities and change the way users act in their environment. For example, firefighters can augment their reality with heat-vision to support them in emergency situations (Abdelrahman, Knierim, Wozniak, Henze and Schmidt, 2017). Exoskeletons can be used to carry heavy loads, thus improving the user's strength and dexterity beyond physiological limitations (Cao, Ling, Zhu, Wang and Wang, 2009; Marcheschi, Salsedo, Fontana and Bergamasco, 2011; Brown, Tsagarakis and Caldwell, 2003). Considering that users often rely on ATs when making decisions, any failure or misjudgment (Stirling, Siu, Jones and Duda, 2018; Bredereke and Lankenau, 2002; Borenstein, Wagner and Howard, 2018) in the joint human-system capabilities could put lives at risk. If we are stronger, smarter, and more perceptive, we can take greater risks. Hence, we expect ATs to increase the risks people take.

Basic and applied research in medicine and psychology has shown that a mere suggestion of improvement, mostly for treatments, can result in real benefits, such as the feeling of being cured of an ailment in response to taking an inert substance (e.g., a sugar pill). This improvement due to a non-specific sham treatment is called the placebo effect (Beecher, 1955; Kaptchuk, 1998), which relies on the individual's expectation that the treatment will improve their condition (Lee and Suhr, 2020).

Studies have shown that placebo effects can also occur when anticipating enhancement (Beedie, Benedetti, Barbiani, Camerone, Lindheimer and Roelands, 2019; Oken, Flegal, Zajdel, Kishiyama, Haas and Peters, 2007; Beedie, M., Colemean and Foad, 2006; Weger and Loughnan, 2013; Rozenkrantz, Mayo, Ilan, Hart, Noy and Alon, 2017). Music (Geers, Weiland, Kosbab, Landry and Helfer, 2005) or medical devices (Magalhães De Saldanha da Gama, Slama, Caspar, Gevers and Cleeremans, 2013; Dawes, Hopkins and Munro, 2013) described as enhancing performance can serve as a placebo. Human-Computer Interaction (HCI) studies showed a placebo effect on system satisfaction for a non-functional social media control system (Vaccaro, Huang, Eslami, Sandvig, Hamilton and Karahalios, 2018) or player behavior with game experience changes for fake elements in a video game (Denisova and Cook, 2019; Denisova and Cairns, 2015). Furthermore, Kosch, Welsh, Chuang and Schmidt (2022) showed altered performance estimates by exposing participants to sham descriptions of a supporting adaptive AI system. Therefore, placebo effects are well-documented for technologies suggesting an improvement. However, research that investigates the placebo effect in ATs is scarce and the anticipation of being augmented may suggest that people are enhanced; hence, they may take more risks when encountering a non-functional system.

This study investigates how performance expectations induced by a sham AT, and thus placebo effects, increase risk-taking. In a user study ($N=30$), participants played a revised version of the Columbia Card Task (Somerville, Haddara, Sasse, Skwara, Moran and Figner, 2019; Weller, King, Figner and Denburg, 2019) assessing risk-taking behavior. The task was adapted, thus we refer to it as RCCT, to briefly show the content of cards, including the location of loss cards, to participants that were shuffled afterwards.

 villa@posthci.com (S. Villa)

 posthci.com (S. Villa)

ORCID(s):

In addition, participants were equipped with an electroencephalograph (EEG).

Following Kosch et al. (2022), we tested each participant with two conditions, each utilizing the presence or absence of an AT that provides cognitive augmentation. In the cognitive augmentation condition, participants were informed that a generated non-audible playing frequency enhanced their cognitive abilities while playing the RCCT. In the non-augmentation condition, participants were told that the system was turned off and they were no longer augmented. In fact, no sound was played during both conditions, and only the narrative description of the AT's status was manipulated. In contrast to Kosch et al. (2022), we were not interested in task performance, but in the influence that the verbal description of the system would have in the user's decision-making, especially risk-taking behavior. Our work demonstrates that descriptions of ATs influence user expectations, the processing of loss information and that anticipation of augmentation increases risk-taking behavior.

2. Background

First, we present investigations on placebo effects drawing from medical studies and psychological studies and then also highlight where they have been researched in the evaluation technology. Next, we provide an overview on ATs and how they are used. Finally, we provide an overview of standard measures of risk-taking to motivate our choice of task in the study.

2.1. The Placebo Effect

Since the 1950s, pioneered by Lasagna, Mosteller, von Felsinger and Beecher (1954), systematic research on placebo mechanisms has been ongoing (Schindel, 2004). A pharmacological placebo effect is the reduction of a symptom, such as pain, due to the patient's own expectations (Price, Finniss and Benedetti, 2008). The placebo effect has been replicated several times in clinical trials (Hróbjartsson and Gøtzsche, 2001) and psychological research (Diederich and Goetz, 2008; Price et al., 2008) and can thus be considered an established research domain.

In addition to the treatment of ailments, researchers found that placebo treatments can improve performance indicators and objective treatment outcomes (Price et al., 2008). Note, however, that placebo effects are most consistently found for subjective evaluations of improvement (Beckham, 1989; Rickels, Hesbacher, Weise, Gray and Feldman, 1970). Numerous studies have demonstrated that the placebo effect can improve a condition, promote creativity, improve cognitive ability, and boost athletic performance (Beedie et al., 2019; Oken et al., 2007; Beedie et al., 2006; Weger and Loughnan, 2013; Rozenkrantz et al., 2017; Watkins, 1986; Belchior, Marsiske, Sisco, Yam, Bavelier, Ball and Mann, 2013). This indicates that a placebo affects the individual's state.

Irrespective of the measurement outcome, the subject's mental model of the effect of treatment is crucial for the placebo effect. Previous research has shown that expectancy

can not only be induced regarding the outcome (response expectancy) (Kirsch, 1999; Oken et al., 2007; Lee and Suhr, 2020), biasing the participant's perception of their reaction (e.g., the placebo alleviating the feeling of pain), but also that evaluation of the stimulus itself can be biased (stimulus expectancy), e.g., the placebo reducing the perception of pain. Implementing response expectancies is the more dominant and robust paradigm (Kirsch, 1999) for eliciting placebo effects; therefore, our sham augmentation will, therefore, focus on the implementation of response expectancies.

Although useful for treatment, the placebo effect obscures treatments' assessment, making placebo control necessary (Price et al., 2008). In placebo-controlled pharmacological studies, participants with impairments are unaware whether they were administered an active (e.g., painkiller) or inactive treatment (e.g., a sugar pill). Therefore, in evaluating these studies, a treatment is considered successful only if its effect supersedes the placebo effect. Likewise, medical standards are used in the evaluation of performance-enhancing drugs to ensure that the deployed treatment is effective (Greely, Sahakian, Harris, Kessler, Gazzaniga, Campbell and Farah, 2008).

In user-centred technology evaluation, systems are probed by a user, typically by comparing an old technology to new technology or standard. In these studies, users are often able to discern the experimental conditions, e.g., a new artificial intelligence (AI) system vs. a baseline system, and identify the system that is intended to provide improvement. Instructions may even indicate the novelty of the tested systems (Caraban, Karapanos, Gonçalves and Campos, 2019) explicitly. Therefore, studies in technology evaluation are susceptible to placebo effects by setting expectations a-priori that affect the interaction and evaluation of systems. For a more nuanced discussion on expectations see e.g., Kujala, Mugge and Miron-Shatz (2017) or biases in technology evaluation, see Caraban et al. (2019) and Kosch et al. (2022).

ATs are rarely compared to a placebo condition, although they are designed to improve human abilities and therefore can elicit expectations of improvement (Cave and Dihal, 2019). Thus, simple verbal descriptions of ATs may produce a placebo effect in the absence of system functionality, thus replicating prior research (Kosch et al., 2022) on human-AI interaction.

While current HCI research has focused on dialogue-oriented interfaces (Shin, 2021a,b) that have been shown to be susceptible to placebo effects (Kosch et al., 2022), we focus on ATs that enhance humans, i.e. interfaces without a dialogue (Rekimoto and Nagao, 1995).

2.2. Human Augmentation Technologies

Parallel to the development of placebo research, Engelbart (1962) has pioneered research on human augmentation (Schmidt, 2017; Engelbart and English, 1968). ATs improve a person's ability to sense, think, or act. For this, ATs typically integrate novel sensors such as thermal cameras, actuators such as vibrotactile motors or exoskeletons, and

include technologies such as AI that supersede human abilities; however, the primary distinction between ATs and these technologies alone is that ATs are embodied by the user and have the goal to enhance human capabilities (Shneiderman, 2022; Raisamo, Rakkolainen, Majaranta, Salminen, Rantala and Farooq, 2019; Villa, Niess, Nakao, Lazar, Schmidt and Machulla, 2023). ATs can be classified as sensory, cognitive, or motor augmentation (Raisamo et al., 2019). For example, initial sensory augmentations emerged from the need to compensate for impaired hearing (Proulx, 2010) or vision (Danilov, Tyler and Kaczmarek, 2008). Motor augmentations were first observed as technologies to compensate for constrained mobility (Kumar, Hote and Jain, 2019). Therefore, early classes of human augmentation have their roots in the development of assistive technologies for assisting users with impairments (Huber, Shilkrot, Maes and Nanayakkara, 2018; Guerrero, da Silva, Fernández-Caballero and Pereira, 2022).

Today, AT researchers are intrigued by the idea of enhancing human abilities beyond evolutionary constraints (Schmidt, 2017; Inami, Uriu, Kashino, Yoshida, Saito, Maekawa and Kitazaki, 2022). For example, Abdelrahman et al. (2017) enabled users for the first time to see the infrared spectrum in a natural way using the combination of augmented reality headsets and thermal cameras, which can support firefighters in operating hot environments (Abdelrahman, Sahami Shirazi, Henze and Schmidt, 2015). Exoskeletons can increase strength for individuals (Cao et al., 2009; Marcheschi et al., 2011; Brown et al., 2003) and enable them to carry heavy loads. Technologies, such as life-logging devices, preserve human memory for longer and with more details than regular memory capacities (Ksibi, Alluhaidan, Salhi and El-Rahman, 2021; Brich, Bause, Hesse and Wesslein, 2019; Dobrowolski, Hanusz, Sobczyk, Skorko and Wiatrow, 2015; Cristina, Jorge, Eva and Mario, 2021). Consequently, extending human capabilities is used not only to re-enable users with impairments but also to enhance users so that they can unlock new potential for interacting with their environment.

AT's mediate the direct interaction of the individual with their surroundings (Rekimoto and Nagao, 1995; Raisamo et al., 2019); augmented senses provide information about the immediate environment via augmented reality or vibrotactile cues, among others (Schmidt, 2017). Motor augmentations, such as exoskeletons, alter the perception of an object's weight, whereas cognitive augmentations modify how individuals interpret or process information. ATs, thus, alter the user's perception of their immediate surroundings and their affordances which may cause the user to behave unexpectedly and engage in risky behavior (Cumiskey, 2017). It has been shown by Low and Chan (2021) that excessive reliance on SCUBA-diving systems, arguably one of the most popular forms of human augmentations, increases the likelihood of risky behavior on the part of the user. Further, Borenstein et al. (2018) illustrated, in the case of exoskeletons, that people trust the exoskeletons' dependability without any prior knowledge of the system and advocate the use

of exoskeletons in high-risk situations, despite the fact that the system was not designed for such situations. In spite of this, it is still possible for people to misjudge the state of the system, even if trust is warranted (Bredereke and Lankenau, 2002); Users can put themselves in physical danger if they are unaware of a system's state and act as if an augmentation is supporting them, such as by pulling a heavy load without the active assistance of an exoskeleton (Stirling et al., 2018). Therefore, expecting enhanced abilities due to augmentation by an AT may increase risk-taking when making decisions.

2.3. Measuring risk-taking

Decision-making describes the process of choosing one option from a set of alternatives (Shafir, Simonson and Tversky, 1993). A decision may be made by calculating risks and benefits or by at least partially relying on emotional responses and gut feelings to each alternative. The latter is known as "hot" decision-making and the former as "cold" decision-making (Buelow and Blaine, 2015). Risk-taking is the decision to take an uncertain action (Deck, Lee, Reyes and Rosen, 2012). Prior research states that individuals who engage in risky behavior do so because they believe that the possible advantages of a given action will outweigh the potential consequences of another (Fromme, Katz and Rivet, 1997). However, risk-taking is not necessarily rational as it is prone to biases that correlate with personality and age and (Nicholson, Soane, Fenton-O'Creevy and Willman, 2005), vary between certain groups, e.g., (Poon, 2016; Hanoch and Gummerum, 2010), and self-assessment e.g., self-perception of skills is linked with higher risk-taking (McKenna and Horswill, 2006).

Such patterns in decision-making under uncertainty can be evaluated using behavioral measurements in lab contexts (Buelow and Blaine, 2015). The most frequent tasks are the Iowa Gambling Task (Buelow and Suhr, 2009), the Balloon Analogue Risk Task (Lejuez, Read, Kahler, Richards, Ramsey, Stuart, Strong and Brown, 2002), and the cold and hot variants (Figner, Mackinlay, Wilkening and Weber, 2009) of the Columbia Card Task (CCT). While all tasks are abstract in nature, they show great external validity (Buelow and Blaine, 2015)

Remarkably they also exhibit internal validity, especially on physiological correlates. In particular, electro-dermal activity, heart rate, functional near-infrared spectroscopy (fNIRS), and EEG responses. Holper and Murphy (2013) reported that participants had stronger activity for electro-dermal activity and fNIRS and decreased heart rate when playing the hot version of the CCT as compared to the cold version, and postulated electro-dermal activity and fNIRS as a suitable combination to study hemodynamic and affective responses of the users.

Event-Related Potentials (ERPs) are time-locked measurements of the EEG activity in response to a particular event or stimulus (Sur and Sinha, 2009). It consists of a series of positive and negative peaks known as components. P300 positivity and N200 negativity (feedback-related negativity, also known as FRN), appearing after 300 ms and

between 200 and 300 ms post-stimulus, respectively, are two particularly important components for stimulus evaluation, selective attention, and conscious discrimination in humans (Patel and Azzam, 2005). Using this as reference, de Groot and Van Strien (2019) demonstrated that feedback evaluation following risky decision-making in the CCT was linked with FRN and a P300 in the EEG, where smaller FRN differences were associated with greater risk-taking in the cold CCT, decreased loss sensitivity, and increased impulsivity, whereas smaller P300 differences were most strongly associated with greater reward responsiveness.

We, therefore, implement and adapt the hot-version of the CCT, as a measure of affective risk-taking in our study and explore recordings of event-related potentials in the EEG.

3. Research model and hypotheses

Previous research has shown that social learning (Kirsch, 1999), classical conditioning (Flaten and Blumenthal, 1999), or verbal information (e.g., expert instructions (Stewart-Williams and Podd, 2004)) can all be used to generate placebos. We concentrated on the latter and attempted to create a placebo effect by changing system descriptions (Kosch et al., 2022). We conducted a within-subjects lab study following medical research on the placebo effect (Gniß, Kappesser and Hermann, 2020) and HCI research on placebo effects (Kosch et al., 2022) to determine the effect of expecting augmentation on risk-taking behavior.

We operationalized risk-taking through the Columbia Card Task (CCT), the participants were given a revised version of the Hot CCT two times, one per condition. In the augmentation condition, participants were told to be supported by a cognitive augmentation. In the no-augmentation condition, participants were told the augmentation system was turned off and that any beneficial effects did not exist. We anticipated that being cognitively augmented affected the participant's risk-taking behavior during the CCT. Note that while trust and over-reliance are relevant to ATs and their adoption, our study focused on the expectation of improvement and the placebo effect's mechanism altering risk-taking.

We used an EEG as a placebo AT and informed participants that the EEG was a Brain-Computer Interface (BCI) playing an inaudible sound that was proven to improve the ability to process information and thus perform better in the CCT. However, the setup was identical for both conditions, the system was not functional, and no sound was played. Note that our study was designed to show the placebo effect of an AT and that we did not focus on BCIs or the efficacy of augmentation of cognitive capacities *per se*, only on the propensity of placebo effects of ATs to increase risk-taking.

We adapted one aspect of the CCT for our study. Typically, in each round of the CCT, a participant is presented with a set of cards face down. Behind these are loss or win cards representing the given amount the player can win or lose. For the purpose of our study, participants were briefly presented with the cards face-up. Afterward, the cards were

put face down again and shuffled using an animation. Participants were led to believe by the verbal description that the augmentation would support them in tracking the cards on the screen. Participants did not know that win or loss cards were rendered at each draw and that the whole game was rigged.

Note that medical research on pain-alleviating placebos has varied a large set of contextual variables that can modulate the placebo effect in size. Wager and Atlas (2015) provides a taxonomy of different contextual cues affecting placebo effects. These cues include the treatment cues (e.g., the novelty of the treatment), the place (e.g., a medical lab), the social situation (e.g., the experimenter wearing a white coat) and verbal suggestion (e.g., describing the mechanism closely); while we have taken these contextual variables into account, our study should resemble a user study for ATs as closely as possible.

To reiterate, we state the following research question:

RQ: Can anticipation of being augmented be induced by verbal description and can this increase risk-taking behavior?

We investigated the following hypotheses to answer this research question:

- H1:** A verbal description of an AT results in an increase in performance expectations
- H2:** A verbal description of an AT results in an increase in performance judgments after interaction
- H3:** Performance expectations improvement induced by a placebo-treatment increases risk-taking behavior.
- H4:** Performance expectations induced by a placebo-treatment affect processing of risk-related information.

4. Method

In the following, we motivate and document our methodological choices in realizing the study. The study implementation with all associated measures can be found at www.anonymizedforanonymity.com.

4.1. Participants

We recruited participants through the university's mailing lists and communication channels. To prevent study participants from detecting the placebo condition (Verbal description of the augmentation system), we refrained from recruiting individuals with prior knowledge of EEG or human augmentation systems. We recruited a total of thirty participants ($N = 30$), one of whom did not consent to the use of their data following the experiment, and two were excluded due to poor data quality (no data was recorded concerning their expectancy ratings). There were a total of twenty-seven participants ($N = 27$, Male = 17, Female = 10, 0 non-binary, 0 participants did not disclose or self-specified a gender) with an average age of 29 years ($M = 29.13$, $SD = 9.51$) and a reported technical competence of ($M = 4.76$, $SD = 1.43$). Participants were compensated 5 euro/30 min for their involvement. The study was approved by an ethics committee (Grant Nr. <removed>).

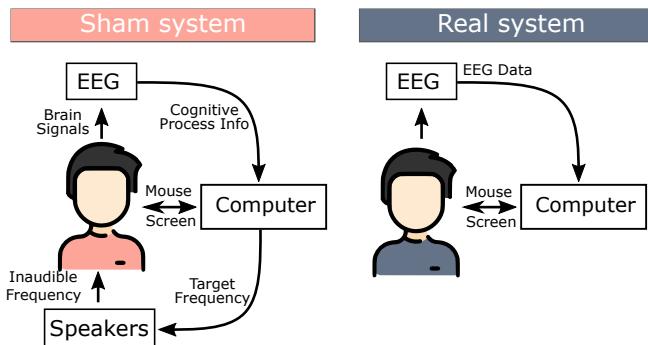


Figure 1: Stimulus: Verbal description; we told participants the EEG was a BCI system that modulated an inaudible sound that improves their information processing and RCCT performance. In reality, the system played no sound and was used to record data only.

4.2. Experimental design

We conducted a within-subjects lab study with four variables of interest, each with two levels. In detail, the independent variables were: (1) Verbal description, referred to as DESCRIPTION (The setup is augmenting participant cognitive skills, referred to as AUGMENTATION condition vs. the setup is not augmenting the participant's cognitive skills, referred to as NO-AUGMENTATION condition), (2) Number of loss cards (one loss card vs. three loss cards) compared to the total number of cards referred to as LOSS CARDS, (3) Value of win cards (10 points vs. 30 points) referred to as WIN AMOUNT, and, (4) Value of loss cards (250 points vs. 750 points) referred to as LOSS AMOUNT. The order of presentation the Verbal description was counter-balanced, while the CCT-related variables (LOSS CARDS, WIN AMOUNT, LOSS AMOUNT) were randomized.

4.3. Stimulus

Verbal description: We compared the influence of two verbal descriptions regarding a human augmentation. We did this by manipulating the system description (i.e., AUGMENTATION condition or NO-AUGMENTATION condition). The participants were informed about the assigned condition before conducting the RCCT (see Figure 1).

During the AUGMENTATION condition, participants were informed that the BCI was analyzing their brain waves to emit an inaudible brain-stimulating sound to boost visual processing, allowing the participants to recognize the win cards more precisely. A coherent explanation of how the system works was provided to the participants. We stated that we used binaural sounds (Colzato, Barone, Sellaro and Hommel, 2017), which are administered through inaudible frequencies (Møller and Pedersen, 2004), are proven to have a positive impact on cognitive functions (e.g., mitigating Alzheimer symptoms Clements-Cortés, Ahonen, Evans, Freedman and Bartel (2016)). In the no-augmentation condition, participants were informed that during this condition, the augmentation device would not be active; therefore, their performance would be determined solely by their ability to visualize the cards shifting, identify the winning cards and

play the game. This condition serves as a control condition in our experiment.

Columbia Card Task related variables According to (Figner et al., 2009), the risk assessment of participants in the Columbia card task is influenced by three variables: the value of win and loss cards, and the number of loss cards in the deck. We used literature informed values for the CCT, namely, 10 and 30 for win cards, 250 and 750 for loss cards, and, 1 and 3 for number of loss cards. The value of win cards is added to the participant's total round score upon flipping a win card. In the same way, the value of loss cards is subtracted from the participant's total score upon flipping a loss card. The number of loss cards in the deck is the number of cards that can lead to point deduction out of the total 27 cards present in the deck.

4.4. Procedure

The participants assignation to the starting condition (augmentation or no-augmentation) was counterbalanced. Participants were supplied with an explanation of the study's design, as well as data protection and comprehensive information. The participants were then requested to grant informed consent to participate in the study in accordance with Declaration of Helsinki and to continue with the demographics and technical competency evaluation. We collected the participants' age, occupation, and identity gender information and seven-point Likert scale ratings of their technical competence.

The researcher then described in full the notion of human augmentation, cognitive augmentation, and the apparatus. Colzato et al. (2017); Møller and Pedersen (2004); Clements-Cortés et al. (2016) works were specifically cited as evidence that the outlined augmentation is functional. However, the augmentation used in our study was a placebo. It was non-functional and did not improve the participants' cognitive abilities. See Figure 2 for an overview.

The induction of placebos adheres to a typical medical research process, see Kosch et al. (2022). Participants received a stimulus consisting of an augmentation system that reportedly enhances human skills and a verbal description. The system was presented as a functional EEG-based human augmentation system that analyzes electric potentials in the brain and boosts performance by playing inaudible sounds to improve cognitive skills, even though no sound is actually created. This design integrates past studies demonstrating that the sound of musical compositions may improve performance through placebo effects (Geers et al., 2005) and that EEG caps can be utilized as a placebo (Magalhães De Saldanha da Gama et al., 2013). The function of the augmentation device in the RCCT was presented as enabling participants to follow the movement of the quickly shuffling cards so they could determine the location of the loss cards, see Figure 2.

The following is an excerpt of the explanation provided to the participants (translated from blinded):

The Placebo Effect of Human Augmentation

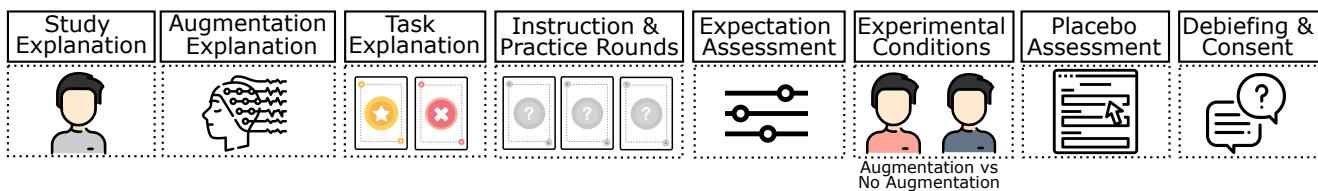


Figure 2: [This figure should be presented in color for print] Study flow diagram: We conducted a within-subjects study. We induced a placebo effect by changing system descriptions. Participants took the Revised Hot Columbia Card Task (RCCT) twice to measure risk-taking. Participants in the AUGMENTATION condition were told that they would be helped by a cognitive augmentation. In the NO-AUGMENTATION condition, participants were told the augmentation system was off and no benefits existed. Finally, we informed them about the actual purpose of the study.

We tune the audio to high and low frequencies that cannot be actively perceived to minimize listener fatigue and distraction from the sounds. For this purpose, the hearing threshold, loudness at which sounds are just heard, is measured. An artificial intelligence (AI) evaluates brain activity during the experiment and dynamically adjusts the binaural tones accordingly. The resulting feedback cycle ensures that the AI optimally adjusts the signal for maximum augmentation and thus maximum performance. In this study, we now want to evaluate whether the system enhances performance and compare this to a control condition without cognitive augmentation by AI.

After describing the augmentation to the participants, we questioned them on their comprehension of the experiment, the augmentation, and its informed purpose (See supplementary material: https://osf.io/gex4t/?view_only=4ed5d5ee9eab40069b908651b6d96a24). This included three questions: *What are the two conditions you will test in this study?*, *How does the augmentation work?*, and, *What are the measured metrics used for?*. Each item had three possible response options, but only one was correct. All participants included in the study answered these questions correctly.

The RCCT was explained once the experimenter checked that the individual understood these points. Participants were informed that their remuneration would depend on their success in each game condition. Thus, they would receive 2.50 euro at the beginning of each card game. The worst scenario would result in 0 euros, while the best outcome would result in 10 euros. They were informed that the actual payout amount would be determined by the number of points obtained at the completion of each condition. At the end of the experiment, all participants were compensated with 5 euro per half hour.

The participants then played two rounds of guided instruction to familiarize themselves with the task. The system guided them through the first round by displaying win and loss cards, and the second round instructed them on how to use the rest of the interface (see Figure 3). After the two instruction rounds, we had the participants play two practice rounds, one of which was intentionally manipulated to demonstrate the risk of flipping loss cards. Following this

explanation and prior to the actual experiment, we did an assessment of performance expectations prior to the RCCT.

Participants underwent a standard auditory threshold detection task across different frequency bands. Thresholds were not of interest in the study but were used to strengthen the placebo system's narrative of the verbal description. Then, depending on the condition (i.e., augmentation or no-augmentation), the participant either receives a pop-up stating that the augmentation is inactive and the game begins, or they are presented with a loading screen where they must wait two minutes until the system allegedly begins generating the inaudible sounds to augment them. After this delay, a message would appear confirming that the augmentation is now active, and then participants would finally be able to play the game. Throughout each condition of the RCCT, we recorded the number of cards flipped and the type of cards flipped. We simultaneously collected EEG data. The conditions were counterbalanced to avoid order effects.

Then we assessed task load and game experience after each condition. After completing both conditions, we measured participant judgments of improvement. Once participants had completed all questionnaires, we examined the usability of the AT, and, finally, debriefed them on the details of the experiments. Then, we measured user judgment of improvement and how they persisted after interaction. After debriefing participants we asked participants if they consented to the use of the collected data once they were fully informed regarding the purpose of the study. The experimenter did not know what their decision was and their decision did not affect their compensation.

4.5. Measures

Assessment of judgments of performance We measured user judgments of performance and how they persisted after interaction. For performance expectations (judgments prior interaction) we used three questions: First, a seven-point Likert item with anchors 1: Strongly disagree, and, 7 Strongly agree compared the expected performance between both conditions: *"I think I will do better in the augmentation condition as compared to the no augmentation condition"*. Then, two slider questions from zero to 930 (the theoretically possible maximum points if the game was not rigged) asking participants the expected number of points in each condition *"How many points do you think you will get in the no augmentation condition in the game?"*, and *"How many points*

The Placebo Effect of Human Augmentation

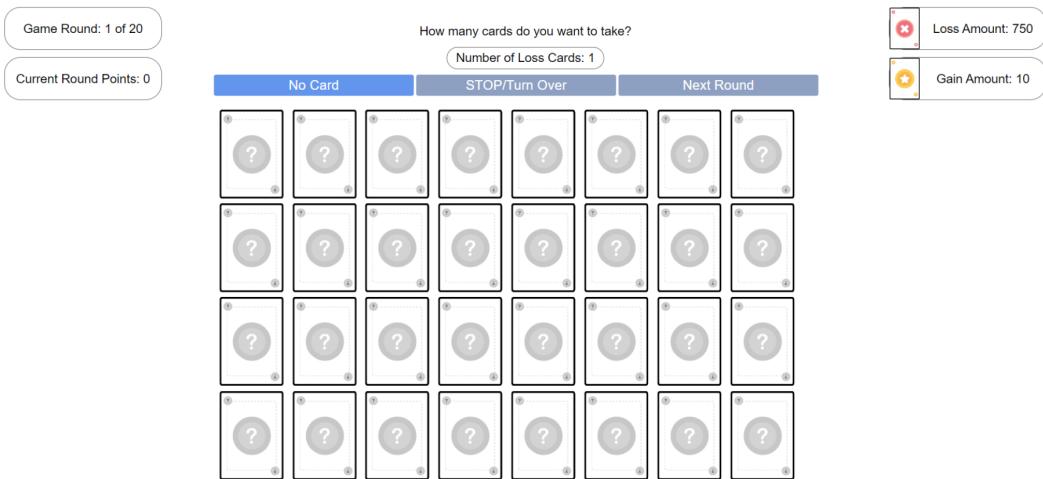


Figure 3: [This figure should be presented in color for print] RCCT interface: The interface was a deck of cards with five indicators: current round, number of points, loss and win card values, and number of loss cards. The interface permits players to skip and stop rounds.

do you think you will get in the augmentation condition in the game?". For judgments of improvement after interaction, we asked participants to rate Table 1 Likert items with anchors 1: Strongly disagree, and, 7 Strongly agree, after completing both conditions.

Risk-taking behavior We applied the Columbia card task (Figner et al., 2009; de Groot and Van Strien, 2019) (hot version) to assess risk-taking behavior. In the CCT, risk-taking is operationalized by the total number of cards flipped in a round by the participants under a set of factors that modulate the risk of flipping a loss card.

There are a predefined number of loss cards (1 vs. 3) in the deck, each of which is equipped with a win (10 vs. 30) and loss point (250 vs. 750). Participants are instructed to flip as many cards as they dare to.

To maintain the task's credibility and prevent participants from simply flipping every card in the deck, seven rounds of each condition game are loss rigged, thus predetermined to result in a loss. These rounds are selected at random. The Columbia card task has been shown to correlate with affective decision-making (Figner et al., 2009; Buelow and Blaine, 2015), risk behavior in adolescents (Panno, 2016), and other experimental measures of risk (de Groot and Van Strien, 2019).

Two variants of the CCT exist (i.e., cold and hot) depending on the ability to interact with the cards in the game. In the hot version, the player must make incremental judgments (i.e., turn over one card at a time) and receive feedback after each decision. In the cold version, the player chooses the number of cards to turn over for the trial. We used the hot version of the CCT for two reasons. First, it measures bias in affective decision-making likely relevant for the use of augmentation under risk. Second, it could be perceived as less random as it allows participants to choose cards individually. The location of loss cards is not known to participants. This means that in the hot version of the

CCT participants can pick cards from arbitrary locations until they encounter a loss-card while in the cold version the algorithm turns over cards sequentially from the beginning. To reiterate, in the cold CCT participants choose the number of cards to turn over while in the hot version of the CCT the participant chooses to flip individual cards.

The verbal description stimulus (DESCRIPTION) goal is to induce the participant's belief to be knowledgeable of the location of loss cards, i.e. the verbal description suggested they have an advantage in selecting the cards due to their enhanced information processing abilities. To allow for participants to know the location of loss cards and thus have an advantage in the game, we adapted the original CCT by showing the location of loss cards briefly using a card flipping animation and then shuffling the cards, referred in this manuscript as RCCT.

In the original CCT, the participant has no visibility of the win and loss cards, so the task depends on the participant's willingness to take risks based on the aforementioned factors (i.e., number of loss cards, amount of gain, amount of loss) that are displayed in the interface. For our narrative, however, we required a skill-based task that is subsequently executed more effectively due to cognitive enhancement. In detail, we implemented two changes (see Figure 4) to the CCT: Each round begins with the deck facing up (one second) so that the player can identify the winning and losing cards, and then the deck is flipped over and shuffled. We repeated the shuffling process five times. The cards are shuffled at an extremely rapid rate. One card could relocate from one side to the other in less than 480ms and its trajectory was shuffled five times before each round, preventing participants from determining the actual location of the cards. The last shuffle lasted 100ms to ensure it was not possible to follow the card location. Thereby preserving the element of risk in the actual task. As in the original CCT the location of loss cards was pre-determined, most rounds were rigged to be

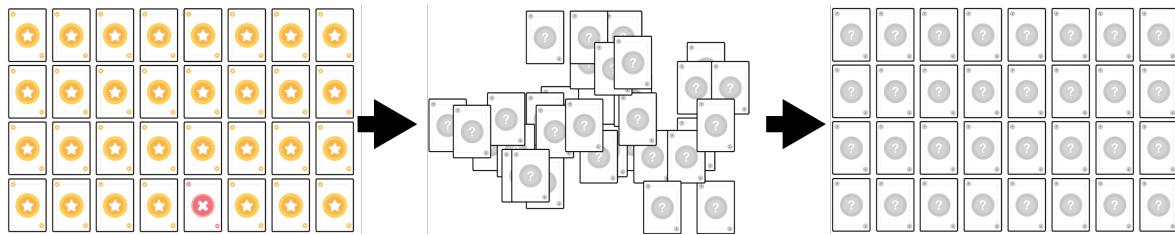


Figure 4: [This figure should be presented in color for print] Revised Columbia Card Task: Each round begins with the deck facing up (one second) so the player can identify the winning and losing cards. The deck is then flipped over and shuffled at an extremely rapid rate and relocated in less than 300ms five times before each round, preventing participants from determining the actual location of the cards and preserving the element of risk in the actual task.

win rounds (13 of 20 rounds; 7 rigged-loss rounds). Thus, only the last or last three cards were loss cards.

Note also that, participants had to decide whether to flip over a card on a given location. Therefore, the augmentation that facilitated the processing of location information was described as giving them a relative advantage in the task. Note that implementing the same routine in the cold CCT would only yield an advantage to participants if they knew all locations of loss cards. Therefore, the hot CCT is better suited for our study.

EEG recordings de Groot and Van Strien (2019) shown that feedback evaluation following risky decision-making in the CCT was linked with feedback-related negativity (FRN) and a P300 in the EEG, where smaller FRN differences were associated with greater risk-taking and, impulsivity, with a decreased loss sensitivity, while smaller P300 differences were most strongly associated with greater reward responsiveness. Therefore we operationalize the processing of risk-related information through the Feedback Related Negativity and P300 in the EEG. For the recording of the EEG, we used an R-Net 64 channel EEG with a wireless amplifier (LiveAmp, Brain Products, Germany) and the corresponding recording software (Brain Vision Recorder) for electrode impedance calibration and the Brain Products LSL Streamer for signal streaming (R-Net, Brain Products, Germany). Electrodes were electrically connected to the scalp using a saline solution. The impedance of the electrodes was kept below $50\text{k}\Omega$ (below the manufacturer's recommendations of $100\text{k}\Omega$). We utilized an average reference and a 500 Hz sampling rate to record the data. We have recorded data from 32 electrodes (see Figure 5).

Task load We distributed a NASA-TLX task load (Hart and Staveland, 1988) questionnaire to compare potential variations in task load generated by the stimulus. It is a widely used subjective assessment tool for evaluating task load. It measures task load by assessing six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. Participants rate each dimension on a scale from 0 to 100, with higher scores indicating higher levels of task load. The NASA-TLX has been extensively validated and is considered a reliable and valid tool for measuring task load in various contexts.

4.6. Apparatus

We used Lab Streaming Layer (LSL) for time-series data acquisition. It was used for networking, time synchronization, and centralized data recording of the EEG streams and the RCCT annotations. We based our RCCT on a web-based CCT experiment provided by The Experiment Factory Sochat (2018) (Stanford, CA). The task was carried out using Microsoft Edge on a Windows (Windows 10 Version 21H2) desktop computer (HP Z1 G6) with an i7 (i7-10700) processor, 16GB of RAM and a screen size of 27 inches with a refresh rate of 60 Hz. Additionally, the web-based experiment was modified to transmit time annotations with the information of each button pressed (i.e., card flip or next round) to the lab streaming layer network and synchronize with EEG data. The participants used the mouse to select the cards in the RCCT. They were positioned in front of the screen, which was calibrated to their eyesight level. The distance between the participant's forehead and the screen was roughly 75 cm (29,5 inch).

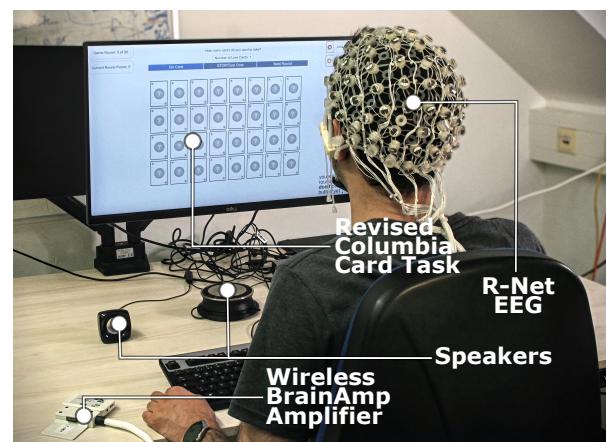


Figure 5: Microsoft Edge was used on a Windows (Windows 10 Version 21H2) desktop computer (HP Z1 G6) with an i7 (i7-10700) processor, 16GB of RAM, and a 27-inch screen with a refresh rate of 60Hz to complete the task. We used an R-Net 64-channel EEG with a wireless amplifier.

4.7. Data analysis

EEG Data Processing To analyze the recorded data, we used the Python MNE library. The data was high pass

filtered at 1 Hz and low pass filtered at 15 Hz (de Cheveigné and Nelken, 2019; Acunzo, MacKenzie and van Rossum, 2012). The data was then re-referenced to the average of all channels, which included the original reference electrode FCz. We applied a notch filter to remove the 50 Hz powerline noise. Then, we sliced the epochs into blocks of -0.3 ms and 0.7 ms, where 0.0 ms denotes the onset of the stimulus. We use the time between -0.3 ms and 0.0 ms as a baseline for the measured stimulus signal. We detected and rejected epochs likely to contain noise using the Autoreject library (Jas, Engemann, Bekhti, Raimondo and Gramfort, 2017). We automatically detected the local maximum around 300ms and 450ms to extract the P300 amplitudes for each epoch according to previous work (de Groot and Van Strien, 2019).

Bayesian Data Analysis and Inference We use a Bayesian approach to data analysis for this paper. We used Bayesian linear mixed models (BLMM). The Bayesian approach has been taken up lately (Kay, Nelson and Hekler, 2016b; Kay, Haroz, Guha and Dragicevic, 2016a; Urbaniak, Tempska, Dowgialło, Ptaszyński, Fortuna, Marcińczuk, Piesiewicz, Leliwa, Soliwoda, Dziublewska, Sulzhitskaya, Karnicka, Skrzek, Karbowska, Brochocki and Wroczyński, 2022; Ackermann, Rusman, Nadolski, Specht and Brand-Gruwel, 2019; Gueron-Sela, Shalev, Gordon-Hacker, Egotubov and Barr, 2023) as it presents several advantages to classical statistics. Kay et al. (2016b) explain advantages of Bayesian statistics in technological contexts that are also relevant to our study. These are in particular: 1. The ability to use prior knowledge and learn from data. 2. To inform on the size of the placebo effect with a given level of precision, 3. It allows for the estimation of effects in small n-studies, 4. The approach enables readers to evaluate the effect size, which can also be close to zero, rather than the mere effect existence.

This Bayesian approach to modeling the CCT is frequently used (Weller et al., 2019; Somerville et al., 2019). Following Weller et al. (2019), we used censoring to model incomplete data distributions (e.g., rigged-loss trials of the CCT). The mean and standard deviation of the data distribution are reported without these censored trials. For a tutorial on Bayesian statistics, a description of the common workflow using brms, and reporting guidelines, see (Schad, Betancourt and Vasishth, 2021; van de Schoot, Depaoli, King, Kramer, Märtens, Tadesse, Vannucci, Gelman, Veen, Willemse et al., 2021; Dix, 2022; Bürkner, 2017). Most importantly, the existence and the non-existence of a placebo effect is likewise important. The Bayesian approach to statistical inference allows us to measure the placebo effect and the non-existence of placebo effects on the measures.

Here, we use Bayesian parameter estimation which allows us to estimate parameter values of effect sizes and quantify the uncertainty regarding these estimates based on the information in our data and the priors applied. We used brms (Bürkner, 2017), a wrapper for the STAN-sampler (Carpenter, Gelman, Hoffman, Lee, Goodrich, Betancourt,

Brubaker, Guo, Li and Riddell, 2017). For statistical inference, we used R (R Core Team, 2021) along with packages for preprocessing (Wickham, Averick, Bryan, Chang, McGowan, François, Grolemund, Hayes, Henry, Hester, Kuhn, Pedersen, Miller, Bache, Müller, Ooms, Robinson, Seidel, Spinu, Takahashi, Vaughan, Wilke, Woo and Yutani, 2019; Bates, Maechler and Jagan, 2022; Müller, 2020; Wickham, Miller and Smith, 2022b; Wickham, Hester and Bryan, 2022a), modeling (Bürkner, 2022, 2017; Bates, Mächler, Bolker and Walker, 2015; Kuznetsova, Brockhoff and Christensen, 2017; Revelle, 2022; Eddelbuettel and François, 2011) and post-processing (Kay, 2022; Groemer, 2020; Makowski, Lüdecke, Ben-Shachar, Patil, Wilson and Wiernik, 2022; Wickham, 2016; Xie, 2015; Chang, 2022; Pedersen, 2020; Lüdecke, Patil, Ben-Shachar, Wiernik, Waggoner and Makowski, 2021; Lüdecke, 2021) the data. We computed 4 Hamilton-Monte-Carlo chains with 40000 iterations each and 10% warm-up samples. Trace plots of the Markov-chain Monte-Carlo permutations were inspected for divergent transitions. All Rubin-Gelman statistics (Gelman, Rubin et al., 1992) were well below 1.1, for effective sampling size.

We compare possible models with approximate leave-one-out cross-validation (LOOCV) (Vehtari, Gelman and Gabry, 2017). This procedure allows us to compare information criteria across models. Relatively smaller LOOCV values indicate a better fit of the model to the data. The best model is then selected and parameters are further analyzed. For these, p_b was computed by calculating the relative proportion of posterior samples being zero or opposite to the median. This metric has similar properties to the classical p -value (Hoijtink and van de Schoot, 2018; Meng et al., 1994; Shi and Yin, 2020) but quantifies the proportion of probability that the effect is zero or opposite given the data observed. Note that this is the reverse of the classical approach to inferential statistics, where one measures the probability of the data given the null-hypothesis with respect to the test statistic. Effects were considered meaningful when there was a particularly low probability ($p_b \leq 2.5\%$) of the effect being zero or the opposite. In addition to the median of the parameter, we calculated the High-Density Interval (HDI) at 95% of the posterior distribution for all parameters, which indicates the possible range of effects given the data, alongside the median of the respective parameter. Simple mean comparisons were done on standardized outcome variables. Therefore, all \tilde{b} represent an effect size in terms of deviations of the standard deviations from the mean (corresponding to Cohen's D for simple effects of categorical predictors with two levels). For models on factorial designs, our analysis of the behavioral and physiological data, we calculated δ_t , which can be interpreted quite similar to Cohen's d and is based on standardizing the population-level effects on the varying-effects and residual variance Judd, Westfall and Kenny (2017); Hedges (2007). We explored the effect of different weakly informative priors on the data. None affected statistical inference. We also provide classical tests resembling Bayesian analysis for each step

of inference and ordinal regression analysis for Likert-type questions in the supplement https://osf.io/gex4t/?view_only=4ed5d5ee9eab40069b908651b6d96a24.

For simple mean comparisons, priors were chosen to resemble only weakly informative priors when standardized with a prior on the standardized mean difference of ($M = 0$, $SD = 1$) and thus encompass positive and negative small to large effect sizes, d_z HDI_{95%} = [-1.96, 1.96], centered at zero on the standardized outcome, for the intercept and the residual a *t*-distributed prior ($df = 3$, $M = 0$, $SD = 1$) was used and we specified a student-link function (v following a γ distribution with $p = .1$, $b = 2$) to resemble the commonly used *t*-test with pooled variances.

5. Findings

We first report on the belief of participants that the system augmented them. Then we analyze user judgment of improvement before and after the stimulus. Followed by modeling risk-taking behavior as a function of Verbal description and judgment of improvement (Kosch et al., 2022). We follow this up with an analysis of feedback-related negativity in response to loss cards (de Groot and Van Strien, 2019) for the EEG signal.

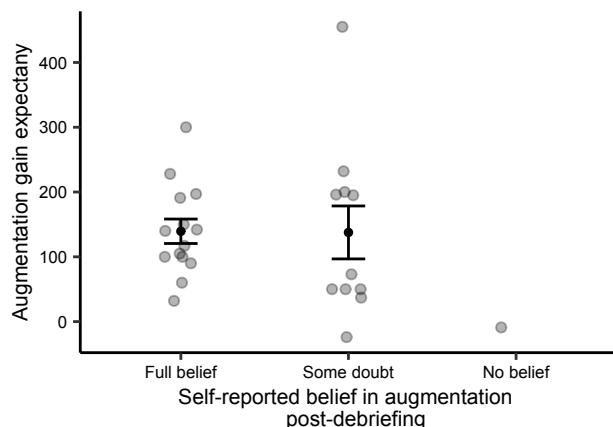


Figure 6: Mean expected augmentation gain in points for the RCCT with individual data points for each subject as a function of self-reported belief in the augmentation after debriefing. Error bars denote ± 1 standard error of the mean.

5.1. Manipulation Check

After the experiment and debriefing participants about the deception and sham treatment, we asked them to indicate whether they believed in the functionality augmentation system or suspected that they were deceived. Only one out of 27 participants (3.70%) indicated that they did not believe in the system's capabilities. Eleven out of 27 (40.74%) participants reported some minor suspicion of the system's functionality (e.g., P2: *I believed that augmentation takes place, but that it really helps was skeptical. I was aware that the difference was more influenced by sequence, fatigue, and other factors.*"). The majority of participants, 14 (51.85%),

fully believed in the augmentation technology's effect. One participant did not disclose whether they believed in the DESCRIPTION or not (3.70%).

5.2. Impact of Verbal Description on Performance Expectations and Judgments of Performance (H1 & H2)

After the description of the experiment, the task, and the system being used but before interaction, we asked participants to indicate how many points they thought they would score with and without the augmentation on a scale ranging from 0 to 930 points. Participants indicated that for the AUGMENTATION condition ($M = 480.30$, $SD = 150.91$), they will score more points as compared to the NO-AUGMENTATION condition ($M = 346.70$, $SD = 130.43$). This difference could be distinguished from zero, $\tilde{b}_{\text{std}} = 0.41$ [0.29, 0.54], $p_b = 0\%$, see Figure 7B. Figure 7A shows the mean for each condition and the substantial variation in participants. While some estimated their gain to be small, others considered it quite substantial. Therefore, hinting at the notion that the placebo effect is subject to high levels of individual variation, which is in line with Kosch et al. (2022).

To inspect whether this variation corresponds to participants' reported judgement of improvement in the AUGMENTATION condition after use, we plotted the difference in expected points, further referred to as relative AUGMENTATION EXPECTANCY (expected points for AUGMENTATION - expected points for NO-AUGMENTATION) as a function of indication of belief on the system (manipulation check). One can see that while on average, there is no substantial difference between the full-belief group and the group of participants that reported some doubt $\tilde{b}_{\text{full/doubt}} = -0.04$ [-0.47, 0.42], $p_b = 56.54\%$, (see Figure 6), the variation is larger in the group that reported some doubt; however the difference in variance between groups was not distinguishable from zero, $\tilde{\sigma}_{\text{full/doubt}} = 0.31$ [-0.01, 0.65], $p_b = 3.12\%$. Also noteworthy is that some participants that voiced minor doubts after the experiment were expecting no gain in points through the augmentation. This is also the case for the one participant who reported that they did not believe in the system at all after the experiment, see Figure 6.

As it could be argued that participants' lack of familiarity with the game mechanics could have impacted our results, we also asked participants to indicate their agreement to "*I think I will do better in the augmentation condition as compared to the no augmentation condition*" on a 7-point Likert scale with anchors 1:Strongly Disagree, and, 7: Strongly Agree. On average, participants reported agreeing with the item with $M = 4.81$ ($SD = 1.47$). We tested this mean against an expected value of 3 (which would indicate neither to agree or disagree with the statement), resembling a one-sample *t*-test. Here, we used a normally-distributed prior on the intercept centered at zero with a *SD* that was two times the standard deviation of the observed variable again with a studentized link-function (v following a γ distribution with $p = .1$, $b = 2$) for the residuals. The sigma prior resembled the mean-comparison model and to allow

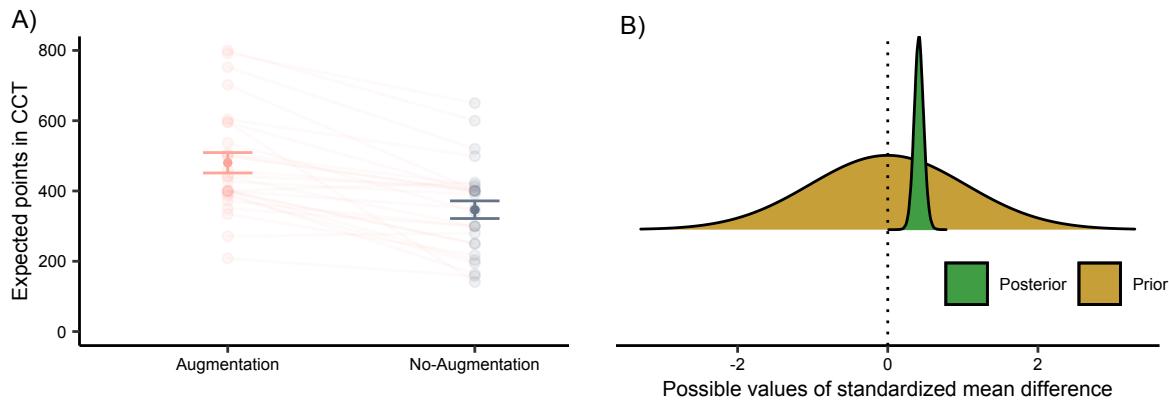


Figure 7: [This figure should be presented in color for print] A: Mean expected points in the RCCT with connected individual mean values. Error bars denote ± 1 standard error of the mean. B: Prior and posterior density plots. Prior samples in beige, and posterior samples in green. The relative density increase from prior to posterior shows how the data has informed the model. No posterior samples lie opposite of zero, indicating that the effect is unlikely to be opposite, or zero.

Table 1

Items were answered on a 7-point likert scale(1 - strongly disagree; 7 - strongly agree). We tested against an indecisive value of 3. Effects that are distinguishable from zero are marked with *. We did not test the SUS against a hypothesized value.

Item/scale	<i>M</i>	<i>SD</i>	\tilde{b}_{std}	<i>HDI</i> _{95%}	<i>p</i> _b
<i>The game was easy to play.*</i>	3.93	1.86	0.47	[0.07, 0.86]	1.07%
<i>The cognitive augmentation has made the task easier.*</i>	3.74	1.40	0.50	[0.12, 0.88]	0.59%
<i>The cognitive augmentation has made the task more enjoyable.</i>	3.56	1.93	0.28	[-0.12, 0.68]	8.24%
<i>The cognitive augmentation has made me more confident.</i>	3.67	1.86	0.35	[-0.05, 0.74]	4.35%
<i>The cognitive augmentation has made me more efficient.</i>	3.52	1.50	0.33	[-0.06, 0.72]	4.63%
<i>The cognitive augmentation has improved my performance.*</i>	4.11	1.55	0.70	[0.31, 1.09]	0.06%
<i>The cognitive augmentation has improved my cognitive abilities.*</i>	3.93	1.54	0.59	[0.21, 0.96]	0.20%
<i>The cognitive augmentation in this game has a lot of potential for future development.*</i>	4.37	1.55	0.85	[0.45, 1.24]	0.00%
System usability scale	56.94	11.34	-	-	-

for more variation a *t*-distributed prior ($df = 3$, $M = 0$, $SD = 1$). The difference between the mean and the expected value of 3 was distinguishable from zero, $\tilde{b}_{\text{std}} = 1.24$ [0.83, 1.61], $p_b = 0.00\%$. We also asked whether they still believed this after interaction with the system and experiencing the NO-AUGMENTATION condition. On average participants still believed in the augmentation, $M = 4.44$, $SD = 1.67$, $\tilde{b}_{\text{std}} = 0.86$ [0.46, 1.24], $p_b = 0\%$) and when comparing their response before and after interaction participants there was no distinguishable reduction in confidence $\tilde{b}_{\text{std}} = -0.12$ [-0.37, 0.14], $p_b = 17.29\%$. The *HDI*_{95%} was centered around zero with a maximum effect of 0.37 *SD* on the outcome variable. Therefore, the belief of superior performance for the AUGMENTATION condition was sustained after interaction, which generated the placebo effect (Kosch et al., 2022).

This placebo effect is also exemplified in the post-experimental questionnaire (see Table 1). We found that participants, on average, judged the augmentation system to facilitate task completion and improve performance and cognitive abilities. This has also prompted participants to

conclude that this augmentation has potential for future development, again see Table 1.

5.3. Influence of Performance expectations on Risk-taking Behavior (H3)

Participants each played 40 rounds of the game. These were sampled from combinations of 2 (1 vs. 3 LOSS CARDS) \times 2 (250 vs. 750 points LOSS AMOUNT) \times 2 (10 vs. 30 points WIN AMOUNT) for each condition of DESCRIPTION. Note that the mixed model approach, we use for analysis does not require equal distribution of trials across experimental variations. For 27 participants, this resulted in 1080 data points that indicated risk-taking as the number of cards turned over in the RCCT. We used censoring for rigged loss rounds to model the whole game in line with Weller et al. (2019). Censoring takes into account that the number of cards in loss rounds only represents a minimum but otherwise unknown estimate of the number of cards the participant would have turned over.

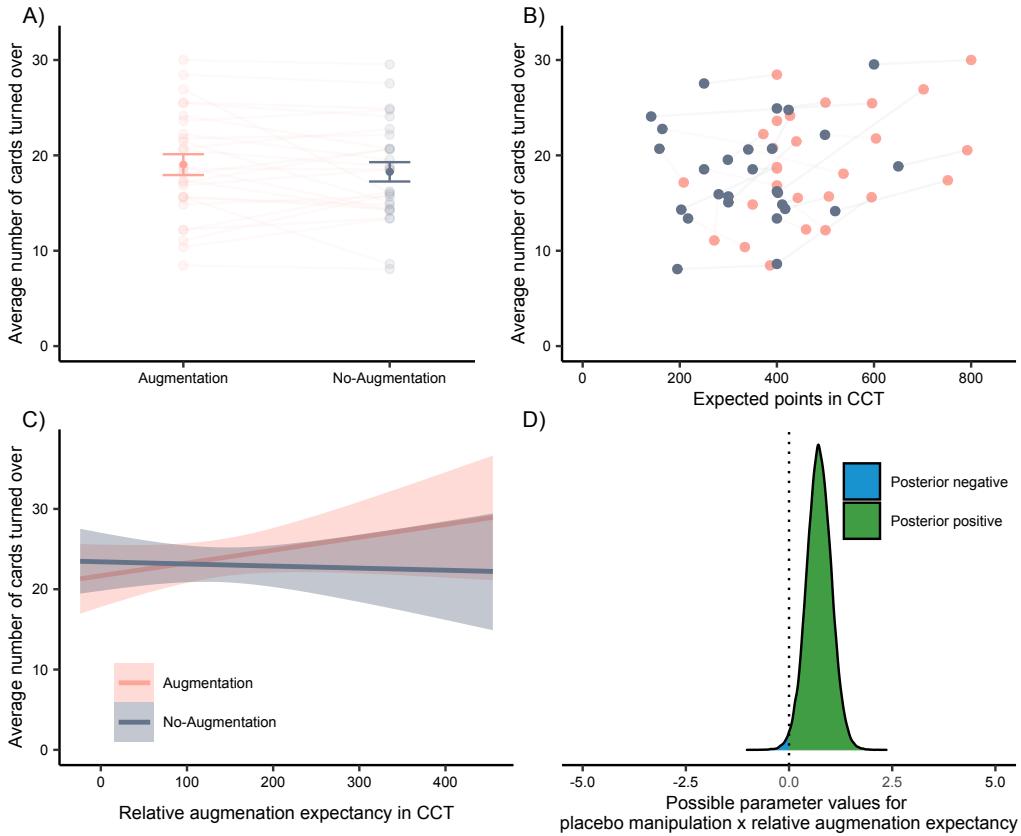


Figure 8: [This figure should be presented in color for print] A: Average number of cards turned over in the RCCT with connected individual mean values. Error bars denote a ± 1 standard error of the mean. B: Average number of cards turned over in the RCCT for each participant as a function of expected points in the RCCT. C: Predicted average number of cards turned over in the RCCT by our model as a relative augmentation gain (Augmentation- No augmentation). D: Posterior density plot. The blue indicates the proportion of posterior samples opposite to the median and thus is a visual representation of the posterior p -value. It quantifies the proportion of probability that the effect is zero or opposite given the data observed. The smaller the blue areas in comparison to the green areas are, the more reliable is the estimation of the effect. We omitted to display the prior distribution as it would appear flat given the wide SD when it is, in fact, normally distributed.

5.3.1. Priors and Model Selection

For multilevel-data and trial-based modelling of the RCCT, we applied normally-distributed priors ($M = 0$, $SD = 10$) on all population-level effects, with Cholesky priors on the unstructured (residual) correlation ($\eta = 2$), and a t -distributed prior ($df = 3$, $M = 0$, $SD = 5$) on the intercept, sigma and the variance, with a normally-distributed prior on the intercept parameters ($M = 20$, $SD = 10$). Two-way interactions in our model were followed up by posterior predictive plots, which serve a similar purpose as post-hoc comparisons in classical statistical inference. We used effect-coding on categorical variables (e.g., 1, -1).

We modeled the effect of the stimulus using a varying intercept for every participant to account for the repeated-measures structure of the data in the mixed model. To allow for individual variation of effects in participants, we added cross-varying slopes for interaction terms for LOSS AMOUNT, WIN AMOUNT, and LOSS CARDS for every subject. The varying intercepts and varying slopes for each

participant serve the purpose of normalization and thus control for systematic individual differences in the dependent variable (e.g., individual differences in loss aversion). All population-level effects of LOSS CARDS, LOSS AMOUNT and WIN AMOUNT, were matched with an interaction term of DESCRIPTION and AUGMENTATION EXPECTANCY (See supplementary material for the full model specification. We compared a null model that only estimated the intercept and the mean (LOO = 4732.99) with a model that accounted for LOSS CARDS, loss amount and win amount with population-level effects and varying-level effects (LOO = 4269.19) similar to Weller et al. (2019) with the LOOCV information criterion and then subsequently added main-effects and fully crossed interaction terms for the DESCRIPTION (LOO = 4218.09) and AUGMENTATION EXPECTANCY (LOO = 4225.04). We selected the most complex model with both DESCRIPTION and AUGMENTATION EXPECTANCY as it allows us to quantify the effect of individual AUGMENTATION EXPECTANCY while providing a fit indistinguishable from

the more parsimonious model. For the sake of brevity, we will analyze the posterior only for this final model.

5.3.2. Posterior Distribution Analysis

As is typically the case for the CCT, our model could show that participants considered the number of LOSS CARDS when making their decision, $\tilde{b}_{\text{loss cards}} = 3.90$ [3.12, 4.69], $p_b = 0.00\%$, $\tilde{\delta}_b = 0.78$ [0.61, 0.94]. They turned over relatively fewer cards ($M = 14.49$, $SD = 6.35$) when there were three LOSS CARDS in the deck as compared to the conditions when there was one LOSS CARDS in the deck ($M = 22.61$, $SD = 4.73$). There was also an effect of LOSS AMOUNT, $\tilde{b}_{\text{loss amount}} = 0.60$ [0.19, 1.01], $p_b = 0.35\%$, $\tilde{\delta}_b = 0.12$ [0.04, 0.20]. With more cards turned in games with 250 points loss possibility ($M = 19.28$, $SD = 4.83$) as compared to 750 points losses ($M = 17.96$, $SD = 5.81$). The WIN AMOUNT did not affect participant's decision to turn over cards, $\tilde{b}_{\text{win amount}} = 0.12$ [-0.42, 0.64], $p_b = 31.53\%$, $\tilde{\delta}_b = 0.02$ [-0.08, 0.13]. Therefore, our data of the RCCT is in line with other psychological studies using the CCT(Somerville et al., 2019; Weller et al., 2019).

We did not find any direct effect of the DESCRIPTION on risk-taking, $\tilde{b}_{\text{description}} = 0.42$ [-0.19, 1.01], $p_b = 8.07\%$, $\tilde{\delta}_b = 0.08$ [-0.04, 0.20]. The HDI indicates that any difference between conditions is smaller than 1 and can therefore be neglected. This lack of a substantial effect, was probably due to the high level of variation in the placebo effect, see Figure 8A. However, we found that relative AUGMENTATION EXPECTANCY (see Figure 8B), increased the number of cards chosen in the AUGMENTATION condition, $\tilde{b}_{\text{description} \times \text{augmentation expectancy}} = 0.72$ [0.12, 1.32], $p_b = 1.04\%$, $\tilde{\delta}_b = 0.15$ [0.02, 0.26], see also Figure 8D. The more participants expected to gain from the AUGMENTATION in the game, the more risks they took when expecting to be augmented, see also Figure Figure 8C. The direct placebo effect term, as well as the interaction effect of DESCRIPTION \times AUGMENTATION EXPECTANCY, were not qualified by any interaction with the factors LOSS AMOUNT, WIN AMOUNT, or LOSS CARDS, all effects centered around zero with $p_b > 15.94\%$. The Bayesian analysis can thus show that relative AUGMENTATION EXPECTANCY is a necessary condition for risk-taking during interaction.

5.4. Task Load

We compared the average NASA TLX Raw sum score across DESCRIPTION. There was no significant difference between conditions, $\tilde{b}_{\text{std}} = -0.03$ [-0.09, 0.04], $p_b = 21.96\%$. Looking closely at the posterior distribution of the mean difference (Figure 9) and taking into account the $HDI_{95\%}$, it is highly unlikely that the AUGMENTATION condition produced any kind of increased subjective workload in the TLX. The $HDI_{95\%}$ indicates that any difference would be smaller than around 1/10 of a point on the sum-score. We can follow that the effect of the DESCRIPTION on the TLX is negligible and not distinguishable from a null-effect. We also found no effect on any of the TLX-subscales, all $p_b > 5.23\%$

5.5. Influence of Performance Expectations on Processing Risk-related Information (H4)

5.5.1. Priors and Model selection

One participant had to be discarded from the dataset due to corrupted data in the recordings. Leaving data from 26 participants for the EEG data analysis. We modeled the EEG separately regarding the amplitude of the FRN and the P300.

For multilevel-data and average-based analysis of the P300 and FRN amplitudes in the EEG, we applied normally-distributed priors ($M = 0$, $SD = 10$) on all population-level effects and varying-level effects, and normally-distributed prior ($M = 0$, $SD = 20$) on the intercept. σ was modeled with a t -distributed prior ($df = 3$, $M = 0$, $SD = 5$) and the student-link function with v following a γ distribution with $p = .1$, $b = 2$. Two-way interactions in our model were followed up by posterior predictive plots, which serve a similar purpose as post-hoc comparisons in classical statistical inference. We used effect-coding on categorical variables (e.g., 1, -1).

To allow for individual variation of win/loss card effects in subjects, we added a varying slope for every subject. The population-level effects of DESCRIPTION, AUGMENTATION EXPECTANCY and win/loss cards were fully crossed (For the full model specification, see supplementary material). As event-related EEG data is prone to outliers, we used a student link function (The deviation of normality was due to the heavy tails of the distribution. For a histogram and a Shapiro-Wilk test). For model selection, we compared a null model that only estimated the intercept, varying slopes, and the mean (LOOFRN = 512.26, LOOP300 = 616.88) with a model that accounted for win/loss cards as population-level effect (LOOFRN = 508.06, LOOP300 = 611.16), and then subsequently added main-effects and fully crossed interaction terms for the DESCRIPTION (LOOFRN = 515.95, LOOP300 = 618.24) and AUGMENTATION EXPECTANCY (LOOFRN = 521.53, LOOP300 = 608.11). For the FRN, the best fit was the NULL model. The LOO information criteria, therefore, suggest that none of the modeled population-level effects had any influence on the amplitude of the FRN. For the P300, the most complex model with all population-level effects had the best fit to the data. We will thus only analyze the posterior of this P300 model.

5.5.2. Posterior distribution analysis

DESCRIPTION affected the strength of the P300, $\tilde{b}_{\text{description}} = -0.76$ [-1.38, -0.08], $p_b = 1.23\%$, $\tilde{\delta}_b = -.24$ [-.47, -0.02]. Participants had higher P300 amplitudes in the NO-AUGMENTATION condition $M = -0.19$, $SD = 5.42$ as compared to the AUGMENTATION condition $M = -0.46$, $SD = 5.24$. We also found a distinguishable effect of AUGMENTATION EXPECTANCY, $\tilde{b}_{\text{expectancy}} = 1.42$ [0.41, 2.38], $p_b = 0.53\%$, $\tilde{\delta}_b = 0.46$ [0.12, 0.82]. With every 10 points of relative AUGMENTATION EXPECTANCY, the amplitude of the P300 increases by about 0.14. There was no distinguishable main effect of win/loss cards on the P300 amplitude $\tilde{b}_{\text{win/loss}} = -1.05$ [-2.13, 0.05], $p_b = 2.97\%$.

The Placebo Effect of Human Augmentation

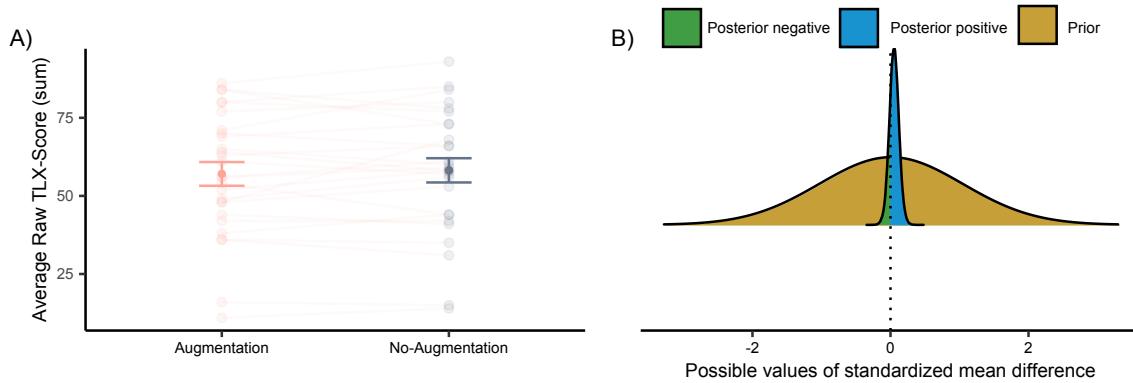


Figure 9: [This figure should be presented in color for print] A: Average TLX sum score with connected individual mean values. Error bars denote ± 1 standard error of the mean. B: Prior and posterior density plots. Prior samples in beige, and posterior samples in green. The relative density increase from prior to posterior shows how the data has informed the model. Posterior samples are centered at zero, indicating that the effect is likely to be small, or zero.

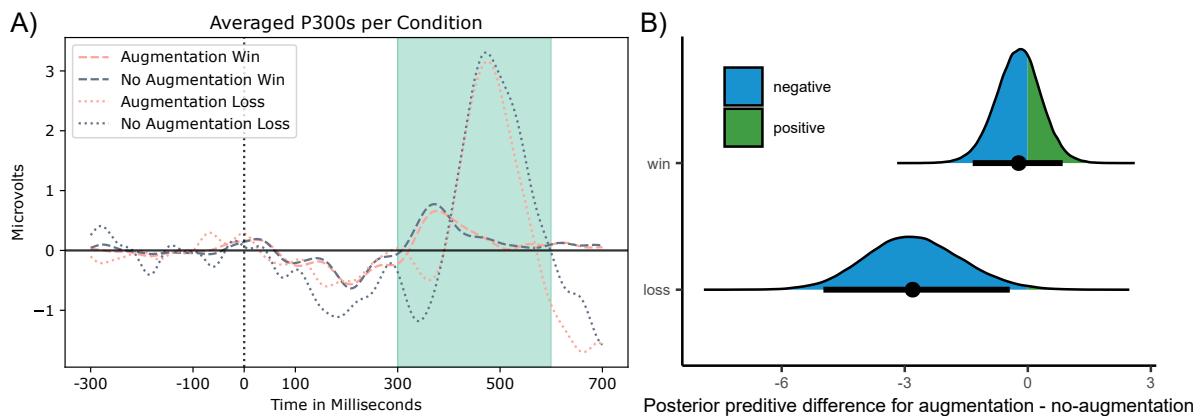


Figure 10: [This figure should be presented in color for print] A: ERP averaged across the central region (Fz, Cz, Pz, Oz). There is a significant decrease in the P300 amplitudes for DESCRIPTION between loss/win -trials for the AUGMENTATION condition. B: Posterior predictive plot for the DESCRIPTION \times win/loss trials interaction.

The DESCRIPTION \times win/loss trials interaction was distinguishable from zero, $\tilde{b}_{\text{description} \times \text{win/loss}} = -0.64$ [-1.23, -0.01], $p_b = 2.31\%$, $\tilde{\delta}_b = -0.21$ [-0.42, 0.00], for posterior predictive plot see Figure 10A. Likewise, a relative AUGMENTATION EXPECTANCY \times win/loss trials interaction was distinguishable from zero, $\tilde{b}_{\text{expectancy} \times \text{win/loss}} = 1.31$ [0.18, 2.39], $p_b = 1.30\%$, $\tilde{\delta}_b = 0.43$ [0.05, 0.81] (see Figure 10B), as well as a DESCRIPTION \times relative AUGMENTATION EXPECTANCY interaction, $\tilde{b}_{\text{description} \times \text{expectancy}} = -0.90$ [-1.45, -0.21], $p_b = 0.85\%$, $\tilde{\delta}_b = -0.31$ [-0.51, -0.09] (see Figure 11A).

Note that these two-way interactions were driven by a the three-way interaction, $\tilde{b}_{\text{description} \times \text{expectancy} \times \text{win/loss}} = -0.97$ [-1.48, -0.32], $p_b = 0.42\%$. To grasp the model estimates and the interaction effects, we compare the raw data to the model predictions Figure 12. One can see that the P300 only increased with AUGMENTATION EXPECTANCY for the NO-AUGMENTATION condition in loss trials; for win cards and loss cards in the NO-AUGMENTATION condition, this correlation was not present. We can thus follow that heightened AUGMENTATION EXPECTANCY is associated with a decreased P300 response for loss trials.

6. Discussion

Our study investigated the placebo effect of ATs and their consequences for risk-taking. We replicated prior research on placebo effects in technology evaluation inducing expectations with a verbal description of an AT (H1) and after using the sham AT, participants maintained their judgment of improvement (H2). Consequently, using ATs results in an inherent perception of improvement in the subject, a placebo effect. While we have not found a direct effect of the placebo on risk-taking, our Bayesian analysis demonstrates that an expectation of improvement is required for increased risk-taking when being told to be augmented (H3). The P300, which typically occurs in the RCCT for loss trials (de Groot and Van Strien, 2019), was lowered when anticipating support from the augmentation compared to the NO-AUGMENTATION condition (H4).

6.1. Experiencing Benefits from a Sham AT

The placebo effect of ATs extends previous studies on placebo effects that focused on improvement after treatment in medical research and psychology (Beedie et al., 2019;

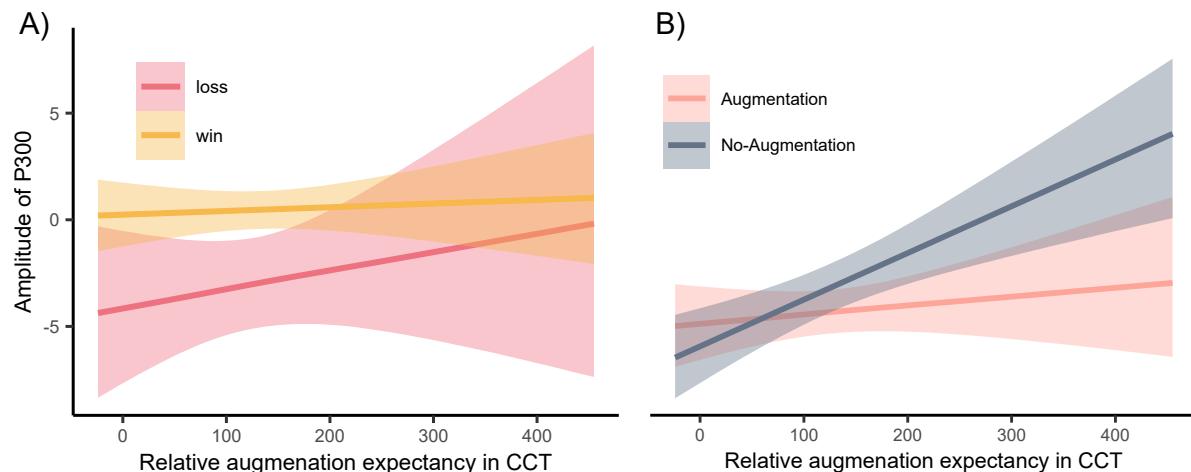


Figure 11: [This figure should be presented in color for print] A: Posterior predictive plot for the win/loss trials \times relative AUGMENTATION EXPECTANCY interaction. B: Posterior predictive plot for the DESCRIPTION \times relative AUGMENTATION EXPECTANCY interaction.

Oken et al., 2007; Beedie et al., 2006; Weger and Loughnan, 2013; Rozenkrantz et al., 2017) but also in technology evaluation (Kosch et al., 2022). In our study, a mere expectation of improvement changed the user's risk-taking, and their expectation of improvement was sustained after use. Particularly interesting is that, in contrast to Kosch et al. (2022), not only the joint performance with the assistance system was increased, but the users' very own capabilities were expected to be improved. Mapping our results onto theories of human-computer integration (Mueller, Lopes, Strohmeier, Ju, Seim, Weigel, Nanayakkara, Obrist, Li, Delfa et al., 2020), our study can assert that perceived human-system capabilities may be judged in the absence of probing system functionality. In this domain of research, our methodology of employing a placebo AT could be used to study how human-system integration affects the users' decision-making. Note, however, that for the use of placebo for research purposes, the mechanisms (Kosch et al., 2022) and contextual variables (Price et al., 2008) in the placebo effect of ATs need to be examined more closely.

6.2. Taking Risks with ATs

Augmentation technologies are mediators of interaction with the real world. Our findings indicate that a belief of being augmented, in conjunction with the user's expectations regarding the AT's performance, is sufficient to modify the user's risk-taking behavior. This must be examined from two standpoints. Firstly, it could be that users pose a risk to themselves. Secondly, the user could engage in risky behavior and endanger those around him. This may be exaggerated in situations where enhancements support in interacting with environments that pose conditions that can not be met with the users' capabilities alone, but only when augmented, e.g., (Abdelrahman et al., 2015, 2017; Borenstein et al., 2018).

Our findings suggest that in these situations, decision-making will be biased in favor of riskier options that match the subjective capabilities of augmentation rather than the objective capabilities of the AT user (Borenstein et al.,

2018). An immediate possibility to prevent placebo effects from promoting risky decision-making would be to support the user in building appropriate mental models about the AT, e.g. by training them to know about the constraints and limitations of the AT. A more advanced strategy would be to support the user in an appropriate control. Here, one could give feedback to the user that human-system capabilities are not enough to meet the user's expectations and therefore foster risk-averse decisions. For this, users' expectations in a given context could be measured verbally (i.e., by polling expectations), extracted from simulated behavior as in the RCCT, or based on physiological sensing (e.g., comparing the amplitude of the P300 for expected and non-expected events). These levels of information could be integrated and presented to the user in an open-loop system. In a closed-loop system, the level of support could be mapped onto expectations in low-risk situations to calibrate the user's mental model, e.g., less support by an exoskeleton when carrying an object that is not too heavy for the user without augmentation. Overall, our study can highlight that decision-making under uncertainty needs to be taken into account when designing ATs, irrespective of the actual human-AT capabilities.

6.3. P300 as a correlate of Risk Processing for ATs

We observed greater P300 amplitudes in the absence of augmentation than in the presence of augmentation for loss-trials. In the context of our study, a reduction in the P300 for loss trials when in the AUGMENTATION condition as compared to the NO-AUGMENTATION condition could have two concurring explanations. First, Gray, Ambady, Lowenthal and Deldin (2004) postulates that in decision-making contexts, the P300 indexes the self-relevance of events. Concerning our study, a reduced P300 for the AUGMENTATION condition could index that non-functional human-AT interaction are processed as less self-relevant. Secondly, one can argue that this was only due to a difference in brain-related potentials caused by perceived ambiguity in

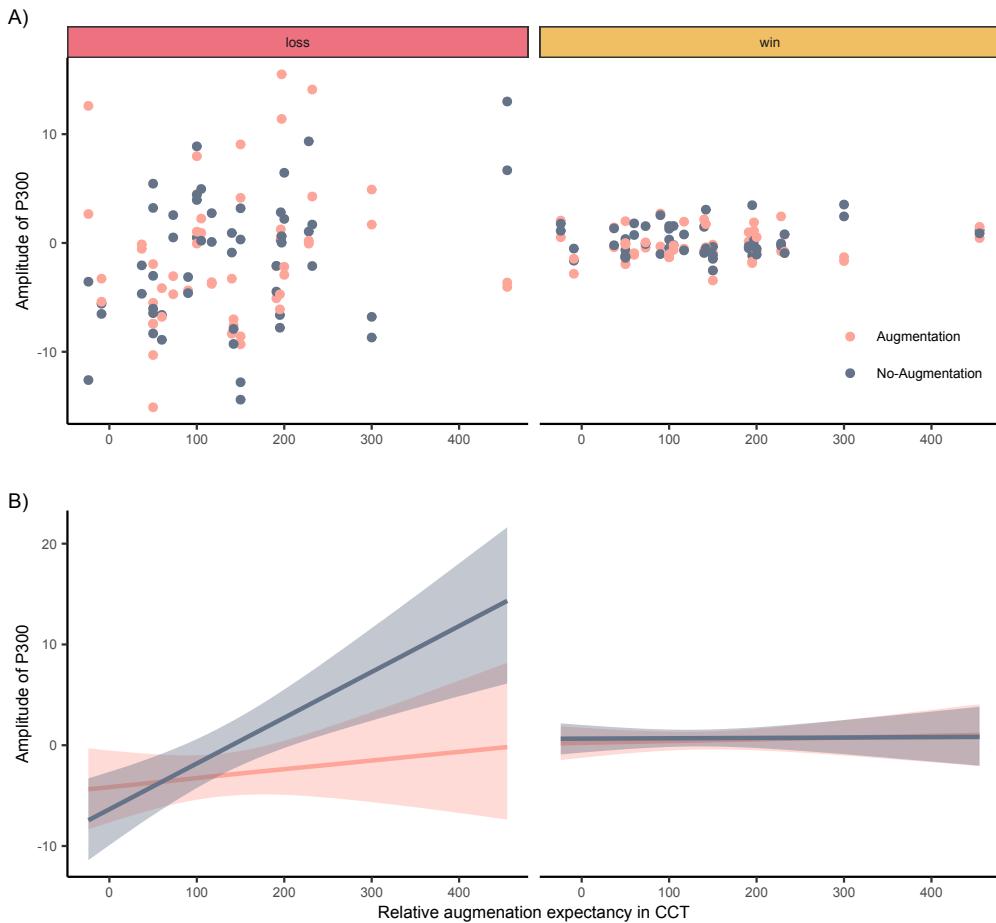


Figure 12: [This figure should be presented in color for print] A: Average P300 for each participant as a function of AUGMENTATION EXPECTANCY and win/loss cards. B: Predicted P300 by our model contrasting AUGMENTATION EXPECTANCY (Augmentation- No augmentation) and win/loss cards.

decision-making. Previous research by Wang, Zheng, Huang and Sun (2015) shows that the P300 amplitude is attenuated in ambiguous situations of risk-taking and less attenuated when there is less ambiguity concerning outcomes. In our study, the AUGMENTATION condition represents less ambiguity as compared to the NO-AUGMENTATION condition because participants subjectively experienced more control over the outcomes of their decisions, i.e., an advantage in knowing where the loss cards are. However, looking closely at our results, the reduced P300 was only found in loss trials and not in win trials and not only as a main effect. Thus, it is likely that self-relevance, as posed by Gray et al. (2004) can explain the pattern in our data. Information about loss trials was not preferably processed as self-relevant when being augmented.

6.4. Effects of ATs on Information Processing in Augmented Individuals

While previous research has suggested that Augmentation Technologies (ATs) may impact self-perception and behavior (Mueller et al., 2020), empirical evidence has been lacking until now; Our results show a notable change in P300

amplitude based on expectancy of augmentation, which may be explained by self-relevance; this finding raises questions about how people process information in tasks performed with AT support. This highlights the significance of developing more effective ATs, given their potential impact on decision-making, as well as the importance of further investigating decision-making when using ATs.

6.5. Evaluation of Human Augmentation

Our study reaffirms (Kosch et al., 2022) position to add a new subcategory of placebos to placebo research, namely, those introduced by digital artifacts.

There are two concurrent processes for describing placebo effects in placebo studies. Expectancy-oriented theories suggest that the occurrence of placebo effects is caused by a rise in treatment efficacy beliefs. Contrarily, conditioned response explanations define the placebo effect in terms of the strength of previously established stimulus-response linkages (such as those between taking a drug and feeling better). As ATs are completely new to individuals, no linkages between stimulus and reaction could have formed. Our findings, therefore, align with the expectancy-based

mechanisms of the placebo effect. Still, one could argue that high-level linkages of novel technology and subjective improvement could have formed, however, this cannot explain the results we obtained on a physiological level that are specific to the integration of loss-information. Also, other mechanisms for placebo induction exist, such as social learning (Kirsch, 1999) that are plausible to persist for ATs and other technologies, e.g., observing someone else receiving a benefit from using an AT and should be explored.

Based on our replication of placebo effects in ATs, we follow that controlling for placebo effects in AT research, much like in psychological and medical intervention studies, is necessary. However, in contrast to medical and psychological research, participants in user studies of technology are often aware of the novelty of a new technology. They can infer which user group they are in. Therefore, in line with Kosch et al. (2022), we recommend implementing measures of placebo control that align with practices and constraints in the particular area of research. We present five ways of controlling for the placebo effects of ATs. These represent neither an exhaustive list nor a general solution to the problem of placebo control for ATs; While some of these guidelines apply to other domains technology evaluation, every study has to be carefully and individually designed to show an effect above and beyond the placebo effect.

Five ways of addressing the placebo effect in the evaluation of ATs:

1. Present a placebo condition with a non-functional AT and compare it to the functional system – *placebo-control*
2. Control for contextual aspects (Wager and Atlas, 2015) that are known to increase placebo effects – *placebo-reduction*
3. Poll expectations before and after use – *placebo-indicator*
4. Consider indirect measures (e.g., physiological measures) when probing the AT – *placebo-indicator*
5. Assess users' qualitative statements in an interview can highlight a mismatch between expectation – *placebo-indicator*

One could argue that this research paves the way for a variety of follow-up investigations for each new technology. However, medical placebo trials can provide a framework for defining the limits of such follow-up research.

First, only studies that can identify the conditions and mechanisms under which placebo effects occur, the potential consequences of placebo effects are relevant to technology evaluation. Here, there is substantial knowledge in the medical literature to start and replicate effects that generalize across technologies.

Second, AI in human-centered AI, or augmentation technologies are examples of technology that create high expectations in their users. Therefore another constraint is that only technologies that raise high expectations may need placebo control. Third, the placebo effects we found are

small and thus false-positive inferences due to placebo effects may only be relevant for user studies that found small effects in statistical comparisons. Overall, while placebo research must be considered in the evaluation of technology, we have to understand the constraints and mechanisms of placebo effects in the evaluation of technology before discrediting large amounts of prior research.

6.6. Implications for Motor and Sensory augmentations

Expectations regarding the perception of external events are known as stimulus expectancy, whereas expectations regarding our own involuntary reactions to events are known as response expectancy (Kirsch, 2018). An example of a response expectancy would be the belief that a sugar pill will improve response time. In contrast, a placebo that improves target detection concentration could be considered a stimulus anticipation. While both expectancy mechanisms to placebo-effects have been studied in the medical domain, it has not yet been determined how these mechanisms contribute to the evaluation of AI augmentation technologies.

For example, Kosch et al. (2022) employed a response expectancy framing technique, informing participants that the task would be easier to complete. However, augmentation technologies can also generate stimulus expectancy. A placebo in sensory augmentation would be considered a stimulus expectancy, whereas a placebo in motor and cognitive augmentation would be a response expectancy. While response expectancies are considered more stable and robust in producing placebo effects, stimulus expectancies rely on the ambiguity of the stimulus (Kirsch, 2018). As placebo effects for stimulus expectancy can be modulated by stimulus ambiguity and are typically weaker than response expectancies in terms of the placebo effect, future research should investigate whether the likelihood of placebo effects varies between augmentation approaches, i.e., sensory, motor, and cognitive.

6.7. Generalizability to other Technological Contexts

Our study has examined the contextual factors related to Cognitive ATs, which is an emerging and highly anticipated technology. Due to the limited understanding of this technology and the external narratives surrounding it, users may develop high expectations of its capabilities (Cave, Coughlan and Dihal, 2019). Similarly, overhyped technologies such as AI have been found to induce placebo effects and affect user performance. Hence, it can be argued that expectations of technologies are central in the judgment of their performance, thus emphasizing the significance of user perception of the technology over its form factor (which was embodied in our study but was desktop-based in (Kosch et al., 2022)). Thus, researchers should consider controlling for users' expectations of the technologies under investigation to prevent potential biases in evaluations and alterations in user behavior. This, for example, implies that tools such

as the Technology Acceptance Model (King and He, 2006) must account for user expectations.

Currently, the placebo literature in Medicine and Psychology emphasizes the role of physical artifacts (e.g. pills) or psychological treatments (Stewart-Williams and Podd, 2004; de Jong, van Baast, Arntz and Merckelbach, 1996). However, placebo effects can be found for game elements, e.g., power-ups, (Denisova and Cook, 2019; Denisova and Cairns, 2015), control modules in user interfaces (Vaccaro et al., 2018), or when being supported by AI (Kosch et al., 2022). Therefore, our study supports the hypothesis that verbal descriptions of digital-technological artifacts can serve as a placebo.

6.8. Limitations and Future Work

Several limitations have to be taken into account concerning our study. First, we did not assess a functional augmentation system. We only compared the placebo to a control condition. Future research should compare all three conditions: A functional augmentation system, a non-functional placebo system, and a control condition. This will allow the researcher to compare the size of the placebo effect, given the context, task, and AT, to the real benefit that the user receives from the functional AT.

Second, we did not examine the differences caused by the type of augmentation. Our narrative was bound to cognitive augmentation technologies and did not include sensory or motor augmentations. A placebo in sensory ATs would change how a user perceives their environment. Considering that stimulus expectancies are not as robust as response expectancies (Kirsch, 1999), it could be that sensory augmentations are less susceptible to placebo effects than cognitive and motor augmentations. One could compare this by presenting the same sham AT to a participant and framing it once in terms of augmenting sensory abilities and once in terms of augmenting the user's cognitive or motor abilities.

Third, we assessed affective risk-taking using a standardized lab-based task. While the hot CCT shows good external validity and correlates with other lab-based measures of affective risk-taking (Buelow and Suhr, 2009; Somerville et al., 2019; Weller et al., 2019), it does not cover more deliberate decision-making under risk (Buelow and Suhr, 2009) which could be studied with the cold CCT. Still, both CCTs are abstract in nature. Ultimately, to understand the aspects of biases in decision-making with ATs, behavior has to be observed in real-world contexts.

Fourth, one could argue that there is a placebo effect on risk-taking; therefore, ATs will not change how people make decisions. Only participants with heightened expectations increased their risk-taking. Therefore, not the belief in augmentation but the heightened expectations pose the problem of a placebo effect of ATs. We may confront this position with the data on the P300 as a retort. While it is true that for risk-taking to manifest, elevated expectations need to be present, the P300 was reduced irrespective of expectancies. Therefore, while only indirect effects on risk-taking are found, we find a direct effect on information

processing in loss trials. Still, to show the direct effects of placebo on risk-taking, larger samples might be needed to reduce uncertainty in the estimation of parameters. Indeed, as we used Bayesian parameter estimation for the analysis of our data, the posterior of our models can be used as a prior for the studies to come.

Finally, our research focuses on Human Augmentation Technologies, which have specific internal characteristics such as form factor and purpose of use (Raisamo et al., 2019) and external characteristics such as narratives, social perception, and expectations (Villa et al., 2023). These characteristics may affect the generalizability of our findings to other types of technologies that do not share these features. Therefore, researchers should consider assessing the placebo effect of various types of technologies.

7. Conclusion

This work shows that expectations with regard to ATs can increase risk-taking. We report a placebo effect, a belief in ATs functionality induced by a verbal description. This belief was sustained after the interaction. We find that participants take more risks during interaction with the AT when they expect the augmentation to support them. We also find a relative reduction in P300, an index of self-relevance of a stimulus, when participants were supposedly augmented and were encountering risk-related information. Our study suggests that placebo effects are relevant for the use of ATs and they affect decision-making under uncertainty, e.g., in safety-critical environments. Likewise, we suggest that much like other fields of human-related research, such as psychology and medicine, research on ATs should consider placebo control. Improvement by an AT must surpass a placebo to constitute a significant improvement.

8. Conflicts of Interest

The authors do not declare any conflict of interest with the submitted work.

References

- Abdelrahman, Y., Knierim, P., Wozniak, P.W., Henze, N., Schmidt, A., 2017. See through the fire: evaluating the augmentation of visual perception of firefighters using depth and thermal cameras, in: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, pp. 693–696.
- Abdelrahman, Y., Sahami Shirazi, A., Henze, N., Schmidt, A., 2015. Investigation of material properties for thermal imaging-based interaction, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 15–18.
- Ackermans, K., Rusman, E., Nadolski, R., Specht, M., Brand-Gruwel, S., 2019. Video-or text-based rubrics: What is most effective for mental model growth of complex skills within formative assessment in secondary schools? *Computers in Human Behavior* 101, 248–258. URL: <https://www.sciencedirect.com/science/article/pii/S0747563219302547>, doi:<https://doi.org/10.1016/j.chb.2019.07.011>.
- Acunzo, D.J., MacKenzie, G., van Rossum, M.C., 2012. Systematic biases in early erp and erf components as a result of high-pass filtering. *Journal of Neuroscience Methods* 209, 212–218. URL: <https://www.sciencedirect.com/science/article/pii/S0165027012002361>, doi:<https://doi.org/10.1016/j.jneumeth.2012.06.011>.
- Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67, 1–48. doi:<https://doi.org/10.18637/jss.v067.i01>.
- Bates, D., Maechler, M., Jagan, M., 2022. Matrix: Sparse and Dense Matrix Classes and Methods. URL: <https://CRAN.R-project.org/package=Matrix>. r package version 1.4-1.
- Beckham, E.E., 1989. Improvement after evaluation in psychotherapy of depression: evidence of a placebo effect? *Journal of clinical psychology* 45 6, 945–50.
- Beecher, H.K., 1955. The powerful placebo. *Journal of the American Medical Association* 159, 1602–1606. URL: <https://doi.org/10.1001/jama.1955.02960340022006>, doi:<https://doi.org/10.1001/jama.1955.02960340022006>.
- Beedie, C.J., Benedetti, F., Barbiani, D., Camerone, E., Lindheimer, J., Roelandts, B., 2019. Incorporating methods and findings from neuroscience to better understand placebo and nocebo effects in sport. *European Journal of Sport Science* 20, 313–325.
- Beedie, C.J., M., S.E., Colemean, D.A., Foad, A.J., 2006. Placebo Effects of Caffeine on Cycling Performance. *Medicine & Science in Sports & Exercise* 38, 2159–2164.
- Belchior, P., Marsiske, M., Sisco, S.M., Yam, A., Bavelier, D., Ball, K., Mann, W.C., 2013. Video game training to improve selective visual attention in older adults. *Computers in Human Behavior* 29, 1318–1324. URL: <https://www.sciencedirect.com/science/article/pii/S074756321300037X>, doi:<https://doi.org/10.1016/j.chb.2013.01.034>.
- Borenstein, J., Wagner, A.R., Howard, A., 2018. Overtrust of pediatric health-care robots: a preliminary survey of parent perspectives. *IEEE Robotics & Automation Magazine* 25, 46–54.
- Bredereke, J., Lankenau, A., 2002. A rigorous view of mode confusion, in: International Conference on Computer Safety, Reliability, and Security, Springer. pp. 19–31.
- Brich, I.R., Bause, I.M., Hesse, F.W., Wesslein, A.K., 2019. Working memory affine technological support functions improve decision performance. *Computers in Human Behavior* 92, 238–249. URL: <https://www.sciencedirect.com/science/article/pii/S0747563218305491>, doi:<https://doi.org/10.1016/j.chb.2018.11.014>.
- Brown, M., Tsagarakis, N., Caldwell, D., 2003. Exoskeletons for human force augmentation. *Industrial Robot: An International Journal* 30, 592–602.
- Buelow, M.T., Blaine, A.L., 2015. The assessment of risky decision making: a factor analysis of performance on the iowa gambling task, balloon analogue risk task, and columbia card task. *Psychological assessment* 27 3, 777–85.
- Buelow, M.T., Suhr, J.A., 2009. Construct validity of the iowa gambling task. *Neuropsychology Review* 19, 102–114.
- Bürkner, P.C., 2017. brms: An r package for bayesian multilevel models using stan. *Journal of statistical software* 80, 1–28.
- Bürkner, P.C., 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80, 1–28. doi:<https://doi.org/10.18637/jss.v080.i01>.
- Bürkner, P.C., 2022. brms: Bayesian Regression Models using Stan. URL: <https://CRAN.R-project.org/package=brms>. r package version 2.17.0.
- Cao, H., Ling, Z., Zhu, J., Wang, Y., Wang, W., 2009. Design frame of a leg exoskeleton for load-carrying augmentation, in: 2009 IEEE International Conference on Robotics and Biomimetics (ROBIO), Ieee.
- Caraban, A., Karapanos, E., Gonçalves, D., Campos, P., 2019. 23 ways to nudge: A review of technology-mediated nudging in human-computer interaction, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 1–15. URL: <https://doi.org/10.1145/3290605.3300733>, doi:<https://doi.org/10.1145/3290605.3300733>.
- Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A., 2017. Stan: A probabilistic programming language. *Journal of statistical software* 76.
- Cave, S., Coughlan, K., Dihal, K., 2019. "scary robots": Examining public responses to ai, in: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, Association for Computing Machinery, New York, NY, USA. p. 331–337. URL: <https://doi.org/10.1145/3306618.3314232>, doi:<https://doi.org/10.1145/3306618.3314232>.
- Cave, S., Dihal, K., 2019. Hopes and fears for intelligent machines in fiction and reality. *Nature Machine Intelligence* 1, 74–78.
- Chang, W., 2022. extrafont: Tools for Using Fonts. URL: <https://github.com/wch/extrafont>. r package version 0.18.
- Clements-Cortés, A., Ahonen, H., Evans, M., Freedman, M., Bartel, L.R., 2016. Short-term effects of rhythmic sensory stimulation in alzheimer's disease: An exploratory pilot study. *Journal of Alzheimer's disease : JAD* 52 2, 651–60.
- Colzato, L.S., Barone, H., Sellaro, R., Hommel, B., 2017. More attentional focusing through binaural beats: evidence from the global-local task. *Psychological Research* 81, 271–277.
- Cristina, O.P., Jorge, P.B., Eva, R.L., Mario, A.O., 2021. From wearable to insideable: Is ethical judgment key to the acceptance of human capacity-enhancing intelligent technologies? *Computers in Human Behavior* 114, 106559. URL: <https://www.sciencedirect.com/science/article/pii/S0747563220303095>, doi:<https://doi.org/10.1016/j.chb.2020.106559>.
- Cumiskey, K.M., 2017. Mobile symbiosis: a precursor to public risk-taking behavior?, in: Mobile Communication. Routledge, pp. 27–31.
- Danilov, Y.P., Tyler, M.E., Kaczmarek, K.A., 2008. Vestibular sensory substitution using tongue electrotactile display. Birkhäuser Basel.
- Dawes, P., Hopkins, R., Munro, K.J., 2013. Placebo effects in hearing-aid trials are reliable. *International Journal of Audiology* 52, 472–477.
- de Cheveigné, A., Nelken, I., 2019. Filters: When, why, and how (not) to use them. *Neuron* 102, 280–293. URL: <https://www.sciencedirect.com/science/article/pii/S0896627319301746>, doi:<https://doi.org/10.1016/j.neuron.2019.02.039>.
- Deck, C., Lee, J., Reyes, J., Rosen, C., 2012. Risk-Taking Behavior: An Experimental Analysis of Individuals and Dyads. *Southern Economic Journal* 79, 277–299.
- Denisova, A., Cairns, P., 2015. The placebo effect in digital games: Phantom perception of adaptive artificial intelligence, in: Proceedings of the 2015 annual symposium on computer-human interaction in play, pp. 23–33.
- Denisova, A., Cook, E., 2019. Power-ups in digital games: The rewarding effect of phantom game elements on player experience, in: Proceedings of the Annual Symposium on Computer-Human Interaction in Play, Association for Computing Machinery, New York, NY, USA. p. 161–168. URL: <https://doi.org/10.1145/3311350.3347173>, doi:<https://doi.org/10.1145/3311350.3347173>.
- Diederich, N., Goetz, C.G., 2008. The placebo treatments in neurosciences. *Neurology* 71, 677–684.

The Placebo Effect of Human Augmentation

- Dix, A., 2022. Bayesian statistics. *Bayesian Methods for Interaction and Design*, 81.
- Dobrowolski, P., Hanusz, K., Sobczyk, B., Skorko, M., Wiatrow, A., 2015. Cognitive enhancement in video game players: The role of video game genre. *Computers in Human Behavior* 44, 59–63. URL: <https://www.sciencedirect.com/science/article/pii/S074756321400661X>, doi:<https://doi.org/10.1016/j.chb.2014.11.051>.
- Eddelbuettel, D., François, R., 2011. Rcpp: Seamless R and C++ integration. *Journal of Statistical Software* 40, 1–18. doi:[10.18637/jss.v040.i08](https://doi.org/10.18637/jss.v040.i08).
- Engelbart, D.C., 1962. Augmenting human intellect: A conceptual framework. Menlo Park, CA , 21.
- Engelbart, D.C., English, W.K., 1968. A research center for augmenting human intellect, in: Proceedings of the December 9–11, 1968, fall joint computer conference, part I, pp. 395–410.
- Figner, B., Mackinlay, R.J., Wilkening, F., Weber, E.U., 2009. Affective and deliberative processes in risky choice: age differences in risk taking in the columbia card task. *Journal of experimental psychology. Learning, memory, and cognition* 35 3, 709–30.
- Flaten, M.A., Blumenthal, T.D., 1999. Caffeine-associated stimuli elicit conditioned responses: an experimental model of the placebo effect. *Psychopharmacology* 145, 105–112.
- Fromme, K., Katz, E.C., Rivet, K., 1997. Cognitive Therapy and Research 21, 421–442.
- Magalhães De Saldanha da Gama, P.A., Slama, H., Caspar, E.A., Gevers, W., Cleeremans, A., 2013. Placebo-suggestion modulates conflict resolution in the stroop task. *PloS one* 8, e75701.
- Geers, A.L., Weiland, P.E., Kosbab, K., Landry, S.J., Helfer, S.G., 2005. Goal activation, expectations, and the placebo effect. *Journal of personality and social psychology* 89, 143.
- Gelman, A., Rubin, D.B., et al., 1992. Inference from iterative simulation using multiple sequences. *Statistical science* 7, 457–472.
- Gniß, S., Kappesser, J., Hermann, C., 2020. Placebo effect in children: the role of expectation and learning. *Pain* 161, 1191–1201.
- Gray, H.M., Ambady, N., Lowenthal, W.T., Deldin, P., 2004. P300 as an index of attention to self-relevant stimuli. *Journal of Experimental Social Psychology* 40, 216–224. URL: <https://www.sciencedirect.com/science/article/pii/S0022103103000921>, doi:[https://doi.org/10.1016/S0022-1031\(03\)00092-1](https://doi.org/10.1016/S0022-1031(03)00092-1).
- Greely, H.T., Sahakian, B.J., Harris, J., Kessler, R.C., Gazzaniga, M.S., Campbell, P., Farah, M.J., 2008. Towards responsible use of cognitive-enhancing drugs by the healthy. *Nature* 456, 702–705.
- Gromer, D., 2020. apa: Format Outputs of Statistical Tests According to APA Guidelines. URL: <https://github.com/dgromer/apa.r> package version 0.3.3.
- de Groot, K., Van Strien, J.W., 2019. Event-related potentials in response to feedback following risk-taking in the hot version of the columbia card task. *Psychophysiology* 56, e13390.
- Gueron-Sela, N., Shalev, I., Gordon-Hacker, A., Egutubov, A., Barr, R., 2023. Screen media exposure and behavioral adjustment in early childhood during and after covid-19 home lockdown periods. *Computers in Human Behavior* 140, 107572. URL: <https://www.sciencedirect.com/science/article/pii/S0747563222003922>, doi:<https://doi.org/10.1016/j.chb.2022.107572>.
- Guerrero, G., da Silva, F.J.M., Fernández-Caballero, A., Pereira, A., 2022. Augmented humanity: A systematic mapping review. *Sensors* 22, 514. URL: <http://dx.doi.org/10.3390/s22020514>, doi:[10.3390/s22020514](https://doi.org/10.3390/s22020514).
- Hanoch, Y., Gummerum, M., 2010. A comparison of the risk-taking behaviors of prisoners and non-prisoners. *Journal of Behavioral Decision Making* 24, 431–442.
- Hart, S.G., Staveland, L.E., 1988. Development of nasa-tlx (task load index): Results of empirical and theoretical research, in: Hancock, P.A., Meshkati, N. (Eds.), *Human Mental Workload*. North-Holland, volume 52 of *Advances in Psychology*, pp. 139–183. URL: <http://www.sciencedirect.com/science/article/pii/S0166411508623869>, doi:[https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9).
- Hedges, L.V., 2007. Effect sizes in cluster-randomized designs .
- Hojtink, H., van de Schoot, R., 2018. Testing small variance priors using prior-posterior predictive p values. *Psychological Methods* 23, 561.
- Holper, L., Murphy, R.O., 2013. Hemodynamic and affective correlates assessed during performance on the columbia card task (cct). *Brain Imaging and Behavior* 8, 517–530.
- Hröbjartsson, A., Götzsche, P.C., 2001. Is the placebo powerless? an analysis of clinical trials comparing placebo with no treatment. *The New England journal of medicine* 344 21, 1594–602.
- Huber, J., Shilkrot, R., Maes, P., Nanayakkara, S. (Eds.), 2018. Assistive Augmentation. Springer Singapore.
- Inami, M., Uriu, D., Kashino, Z., Yoshida, S., Saito, H., Maekawa, A., Kitazaki, M., 2022. Cyborgs, human augmentation, cybernetics, and jizai body, in: *Augmented Humans 2022*, Association for Computing Machinery, New York, NY, USA. p. 230–242. URL: <https://doi.org/10.1145/3519391.3519401>, doi:[10.1145/3519391.3519401](https://doi.org/10.1145/3519391.3519401).
- Jas, M., Engemann, D.A., Bekhti, Y., Raimondo, F., Gramfort, A., 2017. Autoreject: Automated artifact rejection for meg and eeg data. *NeuroImage* 159, 417–429. URL: <https://www.sciencedirect.com/science/article/pii/S1053811917305013>, doi:<https://doi.org/10.1016/j.neuroimage.2017.06.030>.
- de Jong, P.J., van Baast, R., Arntz, A., Merckelbach, H., 1996. The placebo effect in pain reduction: the influence of conditioning experiences and response expectancies. *International journal of behavioral medicine* 3, 14–29. doi:https://doi.org/10.1207/s1527558ijbm0301_2.
- Judd, C.M., Westfall, J., Kenny, D.A., 2017. Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual review of psychology* 68, 601–625.
- Kaptchuk, T.J., 1998. Powerful placebo: the dark side of the randomised controlled trial. *The lancet* 351, 1722–1725. doi:[10.1016/s0140-6736\(97\)10111-8](https://doi.org/10.1016/s0140-6736(97)10111-8).
- Kay, M., 2022. tidybayes: Tidy Data and Geoms for Bayesian Models. URL: <https://CRAN.R-project.org/package=tidybayes>. r package version 3.0.2.
- Kay, M., Haroz, S., Guha, S., Dragicevic, P., 2016a. Special interest group on transparent statistics in hci, in: *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 1081–1084.
- Kay, M., Nelson, G.L., Hekler, E.B., 2016b. Researcher-centered design of statistics: Why bayesian statistics better fit the culture and incentives of hci, in: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 4521–4532.
- King, W.R., He, J., 2006. A meta-analysis of the technology acceptance model. *Information & Management* 43, 740–755. URL: <http://www.sciencedirect.com/science/article/pii/S0378720606000528>, doi:<https://doi.org/10.1016/j.im.2006.05.003>.
- Kirsch, I., 1999. Specifying nonspecifics: Psychological. The placebo effect: An interdisciplinary exploration 8, 166.
- Kirsch, I., 2018. Response expectancy and the placebo effect. *International review of neurobiology* 138, 81–93.
- Kosch, T., Welsch, R., Chuang, L., Schmidt, A., 2022. The placebo effect of artificial intelligence in human-computer interaction. *ACM Transactions on Computer-Human Interaction* .
- Ksibi, A., Alluhaidan, A.S.D., Salhi, A., El-Rahman, S.A., 2021. Overview of lifelogging: Current challenges and advances, pp. 62630–62641.
- Kujala, S., Mugge, R., Miron-Shatz, T., 2017. The role of expectations in service evaluation: A longitudinal study of a proximity mobile payment service. *International Journal of Human-Computer Studies* 98, 51–61.
- Kumar, V., Hote, Y.V., Jain, S., 2019. Review of exoskeleton: History, design and control. 2019 3rd International Conference on Recent Developments in Control, Automation & Power Engineering (RDCAPE) , 677–682.
- Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82, 1–26. doi:[10.18637/jss.v082.i13](https://doi.org/10.18637/jss.v082.i13).
- Lasagna, L., Mosteller, F., von Felsinger, J.M., Beecher, H.K., 1954. A study of the placebo response. *The American journal of medicine* 16, 770–779.

The Placebo Effect of Human Augmentation

- Lee, G.J., Suhr, J.A., 2020. Expectancy effects of placebo neurofeedback in adhd treatment seekers: A neuropsychological investigation. *Neuropsychology* 34, 774.
- Lejuez, C.W., Read, J.P., Kahler, C.W., Richards, J.B., Ramsey, S.E., Stuart, G.L., Strong, D.R., Brown, R.A., 2002. Evaluation of a behavioral measure of risk taking: the balloon analogue risk task (bart). *Journal of experimental psychology. Applied* 8 2, 75–84.
- Low, B.K.L., Chan, A.H.S., 2021. Identification of observable risk-taking behaviours among scuba divers, in: Gutierrez, A.M.J., Goonetilleke, R.S., Robielos, R.A.C. (Eds.), *Convergence of Ergonomics and Design*, Springer International Publishing, Cham. pp. 102–111.
- Lüdecke, D., 2021. sjstats: Collection of Convenient Functions for Common Statistical Computations. URL: <https://strengejacke.github.io/sjstats/>. r package version 0.18.1.
- Lüdecke, D., Patil, I., Ben-Shachar, M.S., Wiernik, B.M., Waggoner, P., Makowski, D., 2021. see: An R package for visualizing statistical models. *Journal of Open Source Software* 6, 3393. doi:10.21105/joss.03393.
- Makowski, D., Lüdecke, D., Ben-Shachar, M.S., Patil, I., Wilson, M.D., Wiernik, B.M., 2022. bayestestR: Understand and Describe Bayesian Models and Posterior Distributions. URL: <https://easystats.github.io/bayestestR/>. r package version 0.12.1.
- Marcheschi, S., Salsedo, F., Fontana, M., Bergamasco, M., 2011. Body Extender: Whole body exoskeleton for human power augmentation, in: 2011 IEEE International Conference on Robotics and Automation, Ieee.
- McKenna, F.P., Horswill, M.S., 2006. Risk taking from the participant's perspective: the case of driving and accident risk. *Health psychology : official journal of the Division of Health Psychology, American Psychological Association* 25 2, 163–70.
- Meng, X.L., et al., 1994. Posterior predictive p-values. *The Annals of Statistics* 22, 1142–1160.
- Mueller, F.F., Lopes, P., Strohmeier, P., Ju, W., Seim, C., Weigel, M., Nanayakkara, S., Obrist, M., Li, Z., Delfa, J., et al., 2020. Next steps for human-computer integration, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–15.
- Møller, H., Pedersen, C., 2004. Hearing at low and infrasonic frequencies. *Noise & health* 6, 37–57.
- Müller, K., 2020. here: A Simpler Way to Find Your Files. URL: <https://CRAN.R-project.org/package=here>. r package version 1.0.1.
- Nicholson, N., Soane, E., Fenton-O'Creevy, M., Willman, P., 2005. Personality and domain-specific risk taking. *Journal of Risk Research* 8, 157–176.
- Oken, B.S., Flegal, K., Zajdel, D., Kishiyama, S., Haas, M., Peters, D., 2007. Expectancy effect: Impact of pill administration on cognitive performance in healthy seniors. *Journal of Clinical and Experimental Neuropsychology* 30, 7–17.
- Panno, A., 2016. Trait emotional intelligence is related to risk taking when adolescents make deliberative decisions. *Games* 7. URL: <https://www.mdpi.com/2073-4336/7/3/23>. doi:10.3390/g7030023.
- Patel, S.H., Azzam, P.N., 2005. Characterization of n200 and p300: Selected studies of the event-related potential. *International Journal of Medical Sciences* 2, 147 – 154.
- Pedersen, T.L., 2020. patchwork: The Composer of Plots. URL: <https://CRAN.R-project.org/package=patchwork>. r package version 1.1.1.
- Poon, K., 2016. Understanding Risk-taking Behavior in Bullies, Victims, and Bully Victims Using Cognitive- and Emotion-Focused Approaches. *Frontiers in Psychology* 7.
- Price, D.D., Finniss, D.G., Benedetti, F., 2008. A comprehensive review of the placebo effect: recent advances and current thought. *Annual review of psychology* 59, 565–90.
- Proulx, M.J., 2010. Synthetic synesthesia and sensory substitution. *Consciousness and Cognition* 19, 501–503.
- R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- Raisamo, R., Rakkolainen, I., Majaranta, P., Salminen, K., Rantala, J., Farooq, A., 2019. Human augmentation: Past, present and future. *International Journal of Human-Computer Studies* 131, 131–143.
- Rekimoto, J., Nagao, K., 1995. The world through the computer: Computer augmented interaction with real world environments, in: Proceedings of the 8th Annual ACM Symposium on User Interface and Software Technology, Association for Computing Machinery, New York, NY, USA. p. 29–36. URL: <https://doi.org/10.1145/215585.215639>, doi:10.1145/215585.215639.
- Revelle, W., 2022. psych: Procedures for Psychological, Psychometric, and Personality Research. URL: <https://personality-project.org/r/psych/>. r package version 2.2.5.
- Rickels, K., Hesbacher, P.T., Weise, C.C., Gray, B., Feldman, H.S., 1970. Pills and improvement: A study of placebo response in psychoneurotic outpatients. *Psychopharmacologia* 16, 318–328.
- Rozenkrantz, L., Mayo, A.E., Ilan, T., Hart, Y., Noy, L., Alon, U., 2017. Placebo can enhance creativity. *PloS one* 12, e0182466.
- Schad, D.J., Betancourt, M., Vasishth, S., 2021. Toward a principled bayesian workflow in cognitive science. *Psychological methods* 26, 103.
- Schindel, L., 2004. The placebo dilemma. *European Journal of Clinical Pharmacology* 13, 231–235.
- Schmidt, A., 2017. Augmenting human intellect and amplifying perception and cognition. *IEEE Pervasive Computing* 16, 6–10.
- van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M.G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., et al., 2021. Bayesian statistics and modelling. *Nature Reviews Methods Primers* 1, 1–26.
- Shafir, E., Simonson, I., Tversky, A., 1993. Reason-based choice. *Cognition* 49, 11–36. URL: <https://www.sciencedirect.com/science/article/pii/001002779390034S>, doi:[https://doi.org/10.1016/0010-0277\(93\)90034-S](https://doi.org/10.1016/0010-0277(93)90034-S).
- Shi, H., Yin, G., 2020. Reconnecting p-value and posterior probability under one-and two-sided tests. *The American Statistician* , 1–11.
- Shin, D., 2021a. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable ai. *Int. J. Hum. Comput. Stud.* 146, 102551.
- Shin, D., 2021b. The perception of humanness in conversational journalism: An algorithmic information-processing perspective. *New Media & Society* 24, 2680 – 2704.
- Shneiderman, B., 2022. Human-centered ai: ensuring human control while increasing automation, in: *Proceedings of the 5th Workshop on Human Factors in Hypertext*, Oxford University Press. pp. 1–2.
- Sochat, V., 2018. The experiment factory: Reproducible experiment containers. *The Journal of Open Source Software* 3, 521. doi:10.21105/joss.00521.
- Somerville, L.H., Haddara, N., Sasse, S.F., Skwara, A.C., Moran, J.M., Figner, B., 2019. Dissecting “peer presence” and “decisions” to deepen understanding of peer influence on adolescent risky choice. *Child development* 90, 2086–2103.
- Stewart-Williams, S., Podd, J., 2004. The placebo effect: dissolving the expectancy versus conditioning debate. *Psychological bulletin* 130, 324.
- Stirling, L., Siu, H.C., Jones, E., Duda, K., 2018. Human factors considerations for enabling functional use of ecosystems in operational environments. *IEEE Systems Journal* 13, 1072–1083.
- Sur, S., Sinha, V.K., 2009. Event-related potential: An overview. *Industrial Psychiatry Journal* 18, 70 – 73.
- Urbaniak, R., Tempka, P., Dowgiallo, M., Ptaszynski, M., Fortuna, M., Marciniuk, M., Piesiewicz, J., Leliwa, G., Soliwoda, K., Dziublewska, I., Sulzhtyskaya, N., Karnicka, A., Skrzek, P., Karbowska, P., Brochocki, M., Wroczynski, M., 2022. Namespotting: Username toxicity and actual toxic behavior on reddit. *Computers in Human Behavior* 136, 107371. URL: <https://www.sciencedirect.com/science/article/pii/S0747563222001935>, doi:<https://doi.org/10.1016/j.chb.2022.107371>.
- Vaccaro, K., Huang, D., Eslami, M., Sandvig, C., Hamilton, K., Karahalios, K., 2018. The illusion of control: Placebo effects of control settings, in: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–13.
- Vehtari, A., Gelman, A., Gabry, J., 2017. Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and computing* 27, 1413–1432.

Villa, S., Niess, J., Nakao, T., Lazar, J., Schmidt, A., Machulla, T.K., 2023. Understanding perception of human augmentation: A mixed-method study. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany URL: <https://doi.org/10.1145/3544548.3581485>.

Wager, T.D., Atlas, L.Y., 2015. The neuroscience of placebo effects: connecting context, learning and health. *Nature Reviews Neuroscience* 16, 403–418.

Wang, L., Zheng, J., Huang, S., Sun, H., 2015. P300 and decision making under risk and ambiguity. *Computational Intelligence and Neuroscience* 2015.

Watkins, M.W., 1986. Microcomputer-based math instruction with first-grade students. *Computers in Human Behavior* 2, 71–75. URL: <https://www.sciencedirect.com/science/article/pii/0747563286900233>, doi:[https://doi.org/10.1016/0747-5632\(86\)90023-3](https://doi.org/10.1016/0747-5632(86)90023-3).

Weger, U.W., Loughnan, S., 2013. Rapid communication: Mobilizing unused resources: Using the placebo concept to enhance cognitive performance. *Quarterly Journal of Experimental Psychology* 66, 23–28.

Weller, J.A., King, M.L., Figner, B., Denburg, N.L., 2019. Information use in risky decision making: Do age differences depend on affective context? *Psychology and aging* 34, 1005.

Wickham, H., 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. URL: <https://ggplot2.tidyverse.org>.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Gromelund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4, 1686. doi:[10.21105/joss.01686](https://doi.org/10.21105/joss.01686).

Wickham, H., Hester, J., Bryan, J., 2022a. *readr: Read Rectangular Text Data*. URL: <https://CRAN.R-project.org/package=readr>. r package version 2.1.2.

Wickham, H., Miller, E., Smith, D., 2022b. *haven: Import and Export SPSS, Stata and SAS Files*. URL: <https://CRAN.R-project.org/package=haven>. r package version 2.5.0.

Xie, Y., 2015. *Dynamic Documents with R and knitr*. 2nd ed., Chapman and Hall/CRC, Boca Raton, Florida. URL: <https://yihui.org/knitr/>. isbn 978-1498716963.