Villa, Steeven; Kosch, Thomas; Grelka, Felix; Schmidt, Albrecht; Welsch, Robin

The placebo effect of human augmentation: Anticipating cognitive augmentation increases risk-taking behavior

Published in: Computers in Human Behavior

DOI: 10.1016/j.chb.2023.107787

Published: 01/09/2023

Document Version
Publisher's PDF, also known as Version of record

Published under the following license: CC BY

The placebo effect of human augmentation: Anticipating cognitive augmentation increases risk-taking behavior

Steeven Villa a,*, Thomas Kosch b, Felix Grelka a, Albrecht Schmidt a, Robin Welsch c

a LMU Munich, Frauenlobstrasse 7a, 80337 Munich, Germany
b HU Berlin, Rudower Chaussee 25, 12489 Berlin, Germany
c Aalto University, Konemiehentie 2, 02150 Espoo, 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 remains after using the sham system 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, & Schmidt, 2017). Exoskeletons can be used to carry heavy loads, thus improving the user’s strength and dexterity beyond physiological limitations (Brown, Tsagarakis, & Caldwell, 2003; Cao, Ling, Zhu, Wang, & Wang, 2009; Marcheschi, Salsedo, Fontana, & Bergamasco, 2011). Considering that users often rely on ATs when making decisions, any failure or misjudgment (Borenstein, Wagner, & Howard, 2018; Bredereke & Lankena, 2002; Stirling, Siu, Jones, & Duda, 2018) in the joint human–system capabilities could put lives at risk. If we are stronger, smarter, and more perceptive, we may 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 & Suhr, 2020).

Studies have shown that placebo effects can also occur when anticipating enhancement (Beedie et al., 2019; Beedie, Stuart Elizabeth, Colemean, & Foad, 2006; Oken et al., 2007; Rozenkrantz et al., 2017; Weger & Loughman, 2013). Music (Geers, Weiland, Kosbab, Landry, & Colemean, 2006; Oken et al., 2007; Rozenkrantz et al., 2017; Weger & Loughman, 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 et al., 2018) or player behavior with game experience changes for fake elements in a video game (Denisova & Cairns, 2015; Denisova & Cook, 2019). Furthermore, Kosch, Welsch, Chuang, and Schmidt (2022) showed altered performance estimates by exposing participants to sham descriptions of a supporting adaptive AI system.

Additionally, 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 remains after using the sham system 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.
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 et al., 2019; Weller, King, Figner, & 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. 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 placebo effects 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, Finnnis, & Benedetti, 2008). The placebo effect has been replicated several times in clinical trials (Hróbjartsson & Gotzsche, 2001) and psychological research (Diederich & 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, & 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, 2006; Belchior et al., 2013; Oken et al., 2007; Rozenzkantz et al., 2017; Watkins, 1986; Weger & Loughnan, 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; Lee & Suhr, 2020; Oken et al., 2007), 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 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 et al., 2008).

In user-centered technology evaluation, systems are probed by a user, typically by comparing an old technology to new technology or a standard technology. 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, & 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., Kajala, 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. 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, 2021b) 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 & Nagao, 1995).

2.2. Human augmentation technologies

Parallel to the development of placebo research, Engelbart (1962) has pioneered research on human augmentation (Engelbart & English, 1968; Schmidt, 2017). 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 supersedes 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 (Raisamo et al., 2019; Shneiderman, 2022; Villa et al., 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, & Kaczmarek, 2008). Motor augmentations were first observed as technologies to compensate for constrained mobility (Kumar, Hote, & Jain, 2019). Therefore, early classes of human augmentation have their roots in the development of assistive technologies for assisting users with impairments (Guerrero, da Silva, Fernández-Caballero, & Pereira, 2022; Huber, Shikroot, Maes, & Nannayakkara, 2018).

Today, AT researchers are intrigued by the idea of enhancing human abilities beyond evolutionary constraints (Inami et al., 2022; Schmidt, 2017). For example, Abdelrahman et al. (2017) enabled users to see in 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, & Schmidt, 2015). Exoskeletons can increase strength for individuals (Brown et al., 2003; Cao et al., 2009; Marcheschi et al., 2011) 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 (Bric, Bause, Hesse, & Wesslein, 2019; Cristina, Jorge, Eva, & Maria, 2021; Dobrowolski, Hanusz, Sobczyk, Skorko, & Wiatrow, 2015; Ksibi, Alahuaidan, Salhi, & El-Rahman, 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 (Raisamo et al., 2019; Rekimoto & Nagao, 1995); 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 & Lankena, 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 (Shafrir, Simonson, & 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 & Blaine, 2015). Risk-taking is the decision to take an uncertain action (Deck, Lee, Reyes, & 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, & 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., (Hanoch & Guumerer, 2010; Poon, 2016), and self-assessment e.g., self-perception of skills is linked with higher risk-taking (McKenna & Horswill, 2006).

Such patterns in decision-making under uncertainty can be evaluated using behavioral measurements in lab contexts (Buelow & Blaine, 2015). The most frequent tasks are the Iowa Gambling Task (Buelow & Suhr, 2009), the Balloon Analogue Risk Task (Lejuez et al., 2002), and the cold and hot variants (Fignier, Mackinlay, Wilkening, & Weber, 2009) of the Columbia Card Task (CCT). While these tasks are abstract in nature, they show great external validity (Buelow & 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 & 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 & 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 hot 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 & Blumenthal, 1999), or verbal information, e.g., expert instructions (Stewart-Williams & Podd, 2004), can 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 (Gnilk, Kappesser, & 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 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 BCIIs 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. Afterwards, 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 entire 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 https://github.com/mimuc/PlaceboAugmentation.

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.

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:

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),

Number of loss cards (one loss card vs. three loss cards) compared to the total number of cards referred to as LOSS CARDS,

Value of win cards (10 points vs. 30 points) referred to as WIN AMOUNT, and,

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 Fig. 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, & Hommel, 2017), which are administered through inaudible frequencies (Meller & Pedersen, 2004), are proven to have a positive impact on cognitive functions (e.g., mitigating Alzheimer symptoms Clements-Cortés, Ahonen, Evans, Freedman, & 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, 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 participant-assignment 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 study information. The participants were then requested to grant informed consent to participate in the study in accordance with initial 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. Clements-Cortés et al. (2016), Colzato et al. (2017), Meller and Pedersen (2004) 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 Fig. 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 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.
The RCCT was explained once the experimenter checked that the participants 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 condition. At the completion of each condition, participants were told the augmentation system was off and no benefits existed. Finally, we informed them about the actual purpose of the study.

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 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 CCT (de Groot & Van Strien, 2019; Figner et al., 2009) (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 (Buelow & Blaine, 2015; Figner et al., 2009), risk behavior in adolescents (Panno, 2016), and other experimental measures of risk (de Groot & 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 Fig. 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 480 ms and its trajectory was shuffled five times before each round, preventing participants from determining the actual location of the cards. The last shuffle lasted 100 ms 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 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) have 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 FRN 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

---

**Table 1**

<table>
<thead>
<tr>
<th>Item/scale</th>
<th>M</th>
<th>SD</th>
<th>k_b</th>
<th>(H_1/d_{\text{graph}})</th>
<th>(p_95)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The game was easy to play.*</td>
<td>3.93</td>
<td>1.86</td>
<td>0.47</td>
<td>[0.07, 0.86]</td>
<td>1.07%</td>
</tr>
<tr>
<td>The cognitive augmentation has made the task easier.*</td>
<td>3.74</td>
<td>1.40</td>
<td>0.50</td>
<td>[0.12, 0.88]</td>
<td>0.59%</td>
</tr>
<tr>
<td>The cognitive augmentation has made the task more enjoyable.</td>
<td>3.56</td>
<td>1.93</td>
<td>0.28</td>
<td>[-0.12, 0.68]</td>
<td>8.24%</td>
</tr>
<tr>
<td>The cognitive augmentation has made me more confident.</td>
<td>3.67</td>
<td>1.86</td>
<td>0.35</td>
<td>[-0.05, 0.74]</td>
<td>4.35%</td>
</tr>
<tr>
<td>The cognitive augmentation has made me more efficient.</td>
<td>3.52</td>
<td>1.50</td>
<td>0.33</td>
<td>[-0.06, 0.72]</td>
<td>4.63%</td>
</tr>
<tr>
<td>The cognitive augmentation has improved my performance.*</td>
<td>4.11</td>
<td>1.55</td>
<td>0.70</td>
<td>[0.31, 1.09]</td>
<td>0.06%</td>
</tr>
<tr>
<td>The cognitive augmentation has improved my cognitive abilities.*</td>
<td>3.93</td>
<td>1.54</td>
<td>0.59</td>
<td>[0.21, 0.96]</td>
<td>0.20%</td>
</tr>
<tr>
<td>The cognitive augmentation in this game has a lot of potential for future development.*</td>
<td>4.37</td>
<td>1.55</td>
<td>0.85</td>
<td>[0.45, 1.24]</td>
<td>0.00%</td>
</tr>
<tr>
<td>System usability scale</td>
<td>56.94</td>
<td>11.34</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

---

* 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.
below 50kΩ (below the manufacturer’s recommendations of 100kΩ). We utilized an average reference and a 500 Hz sampling rate to record the data. We have recorded data from 32 electrodes (see Fig. 5).

**Task load.** We distributed a NASA-TLX task load (Hart & 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, 16 GB of RAM, and a 27-inch screen 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).

### 4.7. Data analysis

**EEG data analysis.** 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 (Acunzo, MacKenzie, & van Rossum, 2012; de Cheveigné & Nelken, 2019). 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, & Gramfort, 2017). We automatically detected the local maximum around 300 ms and 450 ms to extract the P300 amplitudes for each epoch according to previous work (de Groot & Van Strien, 2019). We automatically rejected epochs likely to contain noise using the Autoreject library (Jas, Engemann, Bekhti, Raimondo, & Gramfort, 2017). We automatically detected the local maximum around 300 ms and 450 ms to extract the P300 amplitudes for each epoch according to previous work (de Groot & 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 (Ackermans, Rustman, Nadolski, Specht, & Brand-Gruwel, 2019; Gueron-Sela, Shalev, Gordon-Hacker, Egorubov, & Barr, 2023; Kay, Haroz, Guha & DrAGICevic, 2016; Kay, Nelson & Hekler, 2016; Urbaniak et al., 2022) as it presents several advantages to classical statistics. Kay, Nelson et al. (2016) 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 (Somerville et al., 2019; Weller 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 Bürkner (2017a), Dix (2022), Schad, Betancourt, and Vasishth (2021), van de Schoot et al. (2021). 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, 2017a), a wrapper for
the STAN-sampler (Carpenter et al., 2017). For statistical inference, we used R (R Core Team, 2021) along with packages for pre-processing (Bates, Maechler, & Jagan, 2022; Müller, 2020; Wickham et al., 2019; Wickham, Hester & Bryan, 2022; Wickham, Miller & Smith, 2022), modeling (Bates, Mächler, Bolker, & Walker, 2015; Bürkner, 2017b, 2022; Eddelbuettel & François, 2011; Kuznetsova, Brockhoff, & Christensen, 2017; Revelle, 2022) and post-processing (Chang, 2022; Gromer, 2020; Kay, 2022; Lüdecke, 2021; Lüdecke et al., 2021; Makowski et al., 2022; Pedersen, 2020; Wickham, 2016; Xie, 2015) 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, 1992) were well below 1.1, for effective sampling size.

We compare possible models with approximate leave-one-out cross-validation (LOOCV) (Vehtari, Gelman, & 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, the LOO 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 (Hojtink & van de Schoot, 2018; Meng et al., 1994; Shi & Yin, 2020) but quantifies the proportion of effect that is zero or opposite given the observed data. 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 < 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 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 effect of different weakly informative priors on the data. None affected interpreted quite similar to Cohen’s d and is based on standardizing 𝛿 with two levels). For models on factorial designs, our analysis of the effect of different weakly informative priors on the null-hypothesis with respect to the test statistic. Effects were considered meaningful when there was a particularly low probability 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 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 effect of different weakly informative priors on the data. None affected interpreted quite similar to Cohen’s d and is based on standardizing 𝛿 with two levels). For models on factorial designs, our analysis of the effect of different weakly informative priors on the data. None affected interpreted quite similar to Cohen’s d and is based on standardizing 𝛿 with two levels).

For these, the LOO 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 (Hojtink & van de Schoot, 2018; Meng et al., 1994; Shi & Yin, 2020) but quantifies the proportion of effect that is zero or opposite given the data. 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 < 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 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 effect of different weakly informative priors on the data. None affected interpreted quite similar to Cohen’s d and is based on standardizing 𝛿 with two levels). For models on factorial designs, our analysis of the effect of different weakly informative priors on the data. None affected interpreted quite similar to Cohen’s d and is based on standardizing 𝛿 with two levels).

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 FRN in response to loss cards (de Groot & Van Strien, 2019) for the EEG signal.
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 Fig. 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 ($\gamma$ following a $\gamma$ distribution with $\rho = 1$, $b = 2$) for the residuals. The sigma prior resembled the mean-comparison model and to allow 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, $b_{\text{std}} = 1.24$ [0.83, 1.61], $p_{\text{std}} = 0.00\%$. 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, $M = 4.44$, $SD = 1.67$, $b_{\text{std}} = 0.86$ [0.46, 1.24], $p_{\text{std}} = 0\%$ and when comparing their response before and after interaction participants there was no distinguishable reduction in confidence $b_{\text{std}} = -0.12$ [-0.37, 0.14], $p_{\text{std}} = 17.29\%$. 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 $\text{loss cards}$) $\times$ 2 (250 vs. 750 points $\text{loss amount}$) $\times$ 2 (10 vs. 30 points $\text{win amount}$) 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 RCC. 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.

5.3.1. Priors and model selection

For multilevel data and trial-based modeling of the RCC, 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 $\text{loss amount}$, $\text{win amount}$, and $\text{loss cards}$ 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 $\text{loss cards}$, $\text{loss amount}$ and $\text{win amount}$, were matched with an interaction term of description and augmentation expectancy. See the supplementary material for the full model specification: https://osf.io/gex4t/. We compared a null model that only estimated the intercept and the mean ($\text{LOO} = 4732.99$) with a model that accounted for $\text{loss cards}$, $\text{loss amount}$ and $\text{win amount}$ with population-level effects and varying-level effects ($\text{LOO} = 4269.19$) 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 ($\text{LOO} = 4218.09$) and augmentation expectancy ($\text{LOO} = 4225.04$). 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 $\text{loss cards}$ when making their decision, $b_{\text{loss cards}} = 3.90$ [3.12, 4.69], $p_{\text{std}} = 0.00\%$, $\delta_{\text{std}} = 0.78$ [0.61, 0.94]. They turned over relatively fewer cards ($M = 14.49$, $SD = 6.35$) when there were three $\text{loss cards}$ in the deck as compared to the
The direct effect of the augmentation expectancy was probably due to the high level of variation in the placebo effect, see also Fig. 8C. The lack of a substantial effect, 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.

5.4. Task load

We compared the average NASA TLX Raw sum score across description. There was no significant difference between conditions, \( b_{\text{std}} = -0.03 [-0.09, 0.04], p_b = 21.96\% \). Looking closely at the posterior distribution of the mean difference (Fig. 9) and taking into account the \( H D I_{95\%} \), it is highly unlikely that the augmentation condition produced any kind of increased subjective workload in the TLX. The \( H D I_{95\%} \) 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-subscals, 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 EEG data analysis. We modeled the EEG with a normally-distributed prior (Somerville et al., 2019; Weller et al., 2019).

We did not find any direct effect of the description on risk-taking, \( b_{\text{description}} = 0.42 [-0.19, 1.01], p_b = 8.07\%, \delta_b = 0.08 [-0.04, 0.20] \). 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 Fig. 8A. However, we found that relative augmentation expectancy (see Fig. 8B), increased the number of cards chosen in the augmentation condition, \( b_{\text{description} \times \text{augmentation expectancy}} = 0.72 [0.12, 1.32], p_b = 1.04\%, \delta_b = 0.15 [0.02, 0.26], \) see also Fig. 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 Fig. 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.
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 with \(\nu = 2\). 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 function (the deviation of normality was due to the heavy tails of the distribution. For a histogram and a Shapiro–Wilk test).

Likewise, a relative augmentation expectancy \(\times\) win/loss trials interaction was distinguishable from zero, \(b_{\text{description} \times \text{win/loss}} = -0.64 [-1.23, -0.01], \) \(p_b = 2.31\%\). The relative augmentation expectancy \(\times\) win/loss trials interaction was distinguishable from zero, \(b_{\text{description} \times \text{win/loss}} = 1.31[0.18, 2.39],\) \(p_b = 0.43\) [0.05, 0.81] (see Fig. 10B), as well as a description relative augmentation expectancy interaction, \(b_{\text{description} \times \text{expectancy}} = -0.90 [-1.45, -0.21], \) \(p_b = 0.85\%\), \(\delta_b = -0.31 [-0.51, -0.09]\) (see Fig. 10A).

Note that these two-way interactions were driven by a three-way interaction, \(b_{\text{description} \times \text{expectancy} \times \text{win/loss}} = -0.97 [-1.48, -0.32], \) \(p_b = 0.42\%\). To grasp the model estimates and the interaction effects, we compare the raw data to the model predictions Fig. 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 P900 response for loss trials.

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 CCT
for loss trials (de Groot & 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, 2006; Oken et al., 2007; Rozenkrantz et al., 2017; Weger & Loughnan, 2013) 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 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 cannot be met with the users’ capabilities alone, but only when augmented, e.g., (Abdelrahman et al., 2017, 2015; 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 concurrent explanations. First, Gray, Ambady, Lowenthall, 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 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

The placebo literature so far emphasizes either physical artifacts (e.g., pills) or psychological treatment (Stewart-Williams & Podd, 2004). 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. Contrastingly, 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 & 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 expectancies 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, & 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 & 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 (de Jong, van Baast, Arntz, & Merckelbach, 1996; Stewart-Williams & Podd, 2004). However, placebo effects can be found for game elements, e.g., power-ups, (Denisova & Cairns, 2015; Denisova & Cook, 2019), 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 & Suhr, 2009; Somerville et al., 2019; Weller et al., 2019), it does not cover more deliberate decision-making under risk (Buelow & 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 no direct 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 ATs, 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 risk-taking. We report a placebo effect, a belief in ATs functionality as psychology and medicine, research on ATs should consider placebo control. Improvement by an AT must surpass a placebo to constitute a significant improvement.

CRediT authorship contribution statement

Steeven Villa: Conceptualization, Methodology, Software, Validation, Data curation, Writing – original draft, Writing – review & editing, Visualization. Thomas Kosch: Conceptualization, Methodology, Software, Data curation, Writing – review & editing, Visualization. Felix Grekka: Investigation, Writing – review & editing. Albrecht Schmidt: Resources, Writing – review & editing, Supervision, Funding acquisition. Robin Welsch: Conceptualization, Methodology, Data curation, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data is in OSF, and the link can be found in the paper

Acknowledgments

This work was partly funded by the European Research Council (ERC AMPLIFY, no. 683008), and the Federal Ministry of Education and Research of Germany (BMBF Hive, no. 16SV1833).

Ethic approval information

The German Psychological Society (DGPs) and the local institutional ethics board provided ethical approval for the study: Ethical assessment ID: EK-MIS-2020-023.

References


