LUNDS UNIVERSITET

DEPARTMENT of PSYCHOLOGY

Episodic memory placebo: first try

Pietro Rizzo

Master's Thesis (30 hp)
Spring 2021

Supervisor: Zoltán Kekecs

This study draws upon some of my previously submitted work.

This is the work of many.

Supervisor of the project; thinker of bright ideas; paragon of graciousness

Zoltán Kekecs

My companion. Her encouragement and council were essential.

秋

My parents, who supported and advised me

Gigi *Laura*

The one who answered my call in my hour of programming need, promptly and generously, my friend

Kikko

Participants in the 2nd round of words selection

Melissa *秋*

Volunteers for the 3rd round of words selection: participants and reporters of software issues

Alex *Daniel* *Debora*

Elena *Emanuel* *Giulia*

Kalle *Lalla* *Leon*

Lucas *Ludo* *Lynn*

Marta *Nico* *Olof*

Ottavia *Rebecca* *Renato*

Trava *An anonymous friend of Alex*

Another anonymous friend of Alex *A friend of Emanuel, also anonymous*

A friend of Leon whose name shall not be discolosed *A mysterious figure whose path somehow crossed Leon’s*

The not-so-anonymous mother of Victor

Volunteers for the main experiment: strangers to me and yet participants and reporters of software issues

Anonymous	*Anonymous*	*Anonymous*
Anonymous	*Anonymous*	*Anonymous*
Anonymous	*Anonymous*	*Anonymous*
Anonymous	*Anonymous*	*Anonymous*
Anonymous	*Anonymous*	*Anonymous*
Anonymous	*Anonymous*	*Anonymous*
Anonymous	*Anonymous*	*Anonymous*
Anonymous	*Anonymous*	*Anonymous*
Anonymous	*Anonymous*	

Recruiters

Alex	*Emil*	*Emanuel*
Francesco	*Leon*	*Victor*

Those who created and maintain the Surveys system at University College Cork,
which allows each student at the University to email every other student when in need of participants

Unknown people

Those who created and maintain LUBsearch,
through which students at Lund University can find all sorts of publications

Unknown people

The creator of Sci-Hub

Alexandra Elbakyan

All librarians and programmers everywhere who made literature search possible and easy

Unknown people

All authors whose work was a reference for my own

To all the above go my heartfelt thanks.

Literature shows that it is possible to induce placebo effects through classical conditioning. Literature also offers first evidence of a semantic memory retrieval enhancement effect elicited through a manipulation of expectancy, which may or may not fit the name of semantic memory retrieval placebo. The study here reported sought to discover whether a similar effect could be produced for episodic memory via classical conditioning. To that end, an experiment following a within-participant design was devised, wherein participants would hear pairs of words and recall them in a second moment: a baseline performance would be collected in one condition; in another (‘Acquisition’) a subliminal speech hint would sometimes be given, always coupled with a subliminal speech prime; in a later condition that (conditioned) speech prime would still present itself, but no hint would be given. For the semantic memory retrieval placebo to be found, the number of correctly recalled word pairs in that last condition should have been significantly higher than baseline; this hypothesis was itself conditional on the subliminal hint producing a significant increase in recall performance. A convenience sample (N = 21) participated in the study online. The experiment came out inconclusive due to a failure of the experimental manipulation. Nevertheless, the data collected offers several insights in participants’ behaviour and experience, forming a known base from which an improved version of the experiment may be developed.

Keywords: classical conditioning, cognitive placebo, episodic memory, placebo effects, subliminal speech priming

Memory placebo effects: scattered evidence

A tale of two studies

In 2007, Lowery and his colleagues were conducting a study about the effects of priming on academic test performance, when they came across an unexpected result. The participants in their study's control groups, students subliminally primed with neutral words during an allegedly unrelated task, went on to score significantly higher in their exam if previously told the priming procedure had been designed to boost their exam performance. Lowery and colleagues, however, had aimed their inquiry at the effect of priming students with words related to intelligence; therefore this finding, occurred by accident, was lying beyond the intended scope of their investigation. They dutifully noted the anomaly in their Results section, and spent a few sentences on it in their Discussion, before concluding that further research was necessary. As of today, none of the authors appears to have engaged in said research.

It would seem that, six years later, Weber and Loughan had not read those findings by Lowery et al., when they reported their own: had they done so, they would have found them worth citing. For, what their predecessors had found by chance in control groups, Weber and Loughan deliberately sought in their experiment. Participants in that experiment had to sit in front of a computer monitor and complete a general knowledge quiz comprising twenty multiple-choice questions, each admitting four possible answers, only one of which was correct. Before the quiz started, the researchers told half of the participants that, before each question, they would see the correct answer, flashing for a fleeting moment on the monitor, too quick for reading, too fast for conscious awareness… but, they said, not for the subconscious one. These participants were told that the procedure would boost their performance in an upcoming test, and invited to 'just allow their intuition to speak, because on some level they already knew the answer' (p. 25). The other half of the participants, instead, was told they would see a flicker between each question, denoting the beginning of a new trial. What the two researchers *truly* flashed on the screen between questions was, for both groups, gibberish; but that half which had been led to believe they would perform better outperformed the other in the general knowledge quiz that followed.

Weber and Loughan then dubbed this phenomenon ‘academic placebo’, and proceeded to introduce and discuss it in relation to the larger context of placebo literature. It appeared likely, given the data, that an attempt to persuade participants that they had been subjected to an intervention designed to cause them to perform better had, in fact, caused them to perform better, the intervention itself being void of intrinsic efficacy. Would that description not fit a placebo effect? It would be fair to say that Weber and Loughan never raised the possibility that the placebo frame may not perfectly fit the narrative of their own findings. Indeed, their paper’s abstract ends on the confident claim that “the study documents the relevance of the placebo effect outside the medical and therapeutic setting”.

Reasons for caution

Weber and Loughan’s confidence may be premature to share, for at least two reasons. First, their finding may be explained in some other way. Not as a placebo effect, but, for example, as yet another instance in which people who trust their ‘gut feeling’ more display a tendency to make better assessments (e.g., Mikels et al., 2011): such a dynamic, which is powerful enough to determine an increase in accuracy in guesses about future events (Pham et al., 2012), may well be hypothesised to suffice in explaining how participants given reason to trust their feelings were better able to decide correctly, when faced with four possible answers to questions such as “who painted la Guernica?” (Weber and Loughan, 2013, p. 25). In the words of Pham and colleagues, feelings function as “meta-summaries” of a wealth of information the judgment-making individual possesses, both conscious and unconscious; a “privileged window into all we tacitly know” which people can approach more or less closely, depending on how much they rely on their feelings (Pham et al., p. 468). In this view, approaching the window and staring through it would allow for better calls, at least in certain classes of situations. It is worth noting that the manipulation Weber and Loughan employed to alter participants’ expectancy might also have been adequate had their goal been to experimentally increase reliance on feelings in the making of assessments – the goal of Pham and colleagues' manipulation. Although Weber and Loughan's manipulation probably affected expectancy (probably, because the authors do not report any manipulation check), it must have also, arguably, determined an increase in several factors deemed to be positive moderators of reliance of feelings in forming assessments (Greifeneder 2010): context-related salience of feelings (i.e., the degree to which feelings are attended in a context), backward representativeness of feelings (i.e., the degree to which they are perceived to be caused by the object to be judged – the alternative possible answers, in that case), forward representativeness of feelings (i.e., the degree to which they are perceived to be informative about the object to be judged), and context-related relevance of feelings (i.e., the degree to which they are perceived to be informative for judging a particular dimension about the object to be judged). (It is also worth noting that the manipulation employed by Lowery and colleagues appear far less vulnerable to such a criticism, as it was limited to a few laconic lines on a monitor (p.153), informing participants that the ostensibly unrelated perceptual task they had just completed had been "designed to improve test performance").

The second, and deeper, reason why time is not ripe for calling the phenomenon reported by Weber and Loughan a placebo effect is the following: nothing is known of the effect's underlying mechanisms, not at a psychological level, nor at a neural one; no potential personality correlate has been investigated, nor any genetic one; no developmental trajectory has been hypothesised. The known similarities between the newly found effect and those explored in classical placebo paradigms are, so far, limited to a striking parallelism at a behavioural level: an intrinsically inactive treatment was provided, presented as effective, and proved efficacious. Is that all it takes for a phenomenon to qualify as placebo effect? Now, that is an interesting question, to which we shall come back in a very short while.

For now, let us just note that what Weber and Loughan's results *do* provide, when taken together with those of Lowery and colleagues, is first evidence of a semantic memory retrieval enhancement induced via a manipulation of expectancy.

One perspective on placebo effects

A tentative account

But would such an enhancement be a placebo effect? As promised, we are already back to the question of what the defining characteristics of placebo effects may be. Unfortunately, this question is rather difficult, and, as I show below, the experts in the field have not yet settled on one single, definite, precise definition. As one clear explanation of the concept most central to this paper seems only due, I shall risk providing one myself.

Let O be a set of neuro-psychological events (verbal manipulation of expectancy, classical conditioning, social modelling, etc.) which caused a neuro-psychological association to be formed between a set of events P (taking a pill, talking with a psychotherapist, etc.) and a set of bodily and/or psychological events Q (variation in symptoms, perception, etc.), such that Q is predicted to follow P – the prediction being a neural representation, thus having a certain strength (i.e. probability), and not necessitating the awareness of the animal[1] having it. A placebo (or nocebo) effect on Q via O,

[1] 'Animal' and not 'person', because 'person' is a rather blurry concept, void of a clear biological meaning. 'Animal', and not 'human', because placebo effects have been documented in non-human animals since at least 1962 (Herrnstein). Naturally, when the participants are not human, placebo effects can hardly be produced via manipulation of expectancy, at least in the laboratory, where the procedure typically involves a human researcher telling participants that a certain intervention will have a certain effect. A researcher keen on following such a procedure with, for instance, rat participants, may find that rather daunting language barriers stand in the way of science. Of course, untestable hypotheses are not necessarily false, and non-human animals might well experience expectancy-based placebo effects. For all I know, dolphins in the same pod may have all sorts of shared beliefs about the pharmacological properties of different kinds of

then, is the difference between the effect size of Q registered after the occurrence of P, and that which would have been registered after the occurrence of P in the counterfactual world where O had not taken place. The effect is referred to as 'placebo effect' if it benefits the animal in which it is produced; as 'nocebo effect', instead, if the animal is harmed by it. Please note, as a first corollary to the definition, that, as this account encompasses both *'I shall please'* placebo effects and *'I shall harm'* nocebo ones, it effectively defines a superset which includes the two concepts into a single, generic, neutral entity: a family of effects that we may refer to as 'placebo/nocebo effects'. As a second corollary to the definition, I would add that, naturally, given the variety of possible Os, Ps and Qs, a sizeable number of combinations is possible: thus, when one wishes to discuss placebo effects in general, it is always necessary to use the plural form – and caution.

One may well desire a longer, more informative account; but sketching a satisfactory picture of placebo effects is no easy task. The subject has been portrayed from a variety of perspectives, and, while contemporary representations enjoy a pleasant degree of convergence, they also differ from each other. For instance, while Kirsh put all placebo effects under the explanatory brim of his response expectancy theory (2018), Benedetti (2014), instead, places the accent on the multitude of placebo effects in existence, and on how each of many different outcome measures could be affected through multiple different psychobiological pathways; and so, likewise, while Kaptchuk's endeavours are most strongly associated with the possibility of open-label placebos (2018) and the importance of the clinician-patient interaction (Miller and Kaptchuk, 2008), the work of Colloca gave more prominence to the relevance of learning (Colloca and Miller, 2011), whereas that of Jensen highlighted how awareness of intervention is not required for placebo effects to occur (2018). Furthermore, one of the points on which the authorities appear to systematically agree with each other is that of the necessity of further studies in the field, a necessity which arises from the present imperfection of our knowledge

---

algae, and those beliefs could cause an ill one to seek and consume a specific plant, as well as to experience an improvement afterwards.

of the phenomenon. Sum, then, the plurality of views on placebo effects with their admitted incompleteness, and it will be easy to see how a full portrait of the subject is just not feasible at the moment – here, or elsewhere.

Roads and moderators

Now that a rough sketch of the whole beast has been provided, two elements of the drawing must be fleshed out a bit: first, those mechanisms, both psychological and neural, thought to underlie placebo effects (in other words, the 'Os' of the definition above); and, second, those individual factors thought to moderate the emergence of placebo effects, which consist of psychological constructs (of which we shall say more) and genetic differences (of which we shall say *no* more – but consult Hall et al., 2015, and Wang et al., 2017, for an overview).

Let us consider mechanisms first. It is apparent, today, that placebo effects may be reached through different routes. However, the royal road is, undoubtedly, manipulation of expectancy (e.g., Petrie & Rief, 2019); this way leads to positive results even in open-label placebo research (Wei et al., 2018). It is not the only way, though; not by far. Operant conditioning has been shown to lead to placebo effects as well (Adamczyk et al., 2019), like observational learning (Bajcar & Bąbel, 2018); and the quality of clinician-patient interplay has been proven to be of relevance (see, for example, the delightful history of placebo research, authored by Finnis in 2018, and his reflections presented therein on results by Gracely et al., reported in 1985, and Kaptchuk et al., reported in 2008, respectively). Finally, but second only to manipulation of expectancy for prominence, comes classical conditioning, which a most intriguing line of research brought forward by Jensen and colleagues (2012, 2014, 2015; cf. Bąbel, 2019) has shown to also achieve placebo - outside of conscious awareness.

Let us then consider moderators. If mechanisms can be thought of as roads, then moderators may have to do with how willing one is to walk, how deep one wishes to delve inside the placebo effect territory. While over a dozen of psychological constructs have been considered as moderators, a recent systematic review has shown that dispositional optimism as assessed through the Life Orientation Test-Revised (LOT-R; Scheier et al., 1994) is the one most likely to moderate placebo effects, with a positive correlation between optimism and the placebo[2] effect (Kern et al., 2020). Different studies have contributed to establish this link; studies, interestingly, looking at a variety of placebo effects, including analgesia (Geers et al., 2010; Morton et al., 2009), sleep quality improvement (Geers et al., 2007), and stress recovery (Darragh et al., 2014).

Placebo matters

This digression into the nature of placebo effects being concluded, let us re-approach the possibility of a genuine memory placebo effect. Such a phenomenon may or may not exist; should we care? The answer is, of course, yes. Here is why.

The value of understanding placebo effects' fundamental dynamics cannot be emphasized enough – but we shall try. To begin with, the ability of researchers to successfully control for placebo effects is one critical methodological pillar upon which a large part of medical knowledge is built; thus, a shift in our comprehension of those phenomena has the potential to shake many current convictions and simultaneously open new perspectives in a wide range of fields, both theoretical and applied. Furthermore, placebo effects are fascinating in and of themselves, since their counterintuitive nature forces us to consider the mind-body system from new perspectives. Finally, should the reasons of science and curiosity not suffice, consider that placebo effects are (and have arguably been for at least a few hundred millennia, though probably much longer) a key component of possibly every treatment of any kind: they account for significant amount of the favourable outcomes of therapy, both in the medical and psychological fields (Petrie & Rief, 2019). Therefore, a deeper knowledge of

[2] *Strictu sensu.* The relationship between dispositional optimisms and nocebo effects has also been probed, but with discordant outcomes (Kern et al., 2020).

their inner workings may enable a better harnessing of their power, resulting in more effective therapeutic tactics, with minimal or no side-effects, to the benefit of an enormous number of patients, present and future (Benson & Friedman, 1996; Finniss, 2018). Research on placebo effects is then an endeavour of public interest.

Nonetheless, our knowledge of such processes remains inadequate. The rich and vibrant research field devoted to this topic has concentrated most of its attention on placebo effects affecting changes in a limited range of domains, perhaps of more immediate medical relevance; pain perception being the most prominent (Klinger et al., 2018), arguably followed by drug effects, immune response regulation, and motor performance. And here is where the possibility of memory placebo effect comes into play. Because such cognitive placebo effects are investigated, too, but not nearly as much.

Cognitive placebo effects

When one thinks of placebo studies in basic research, the most typical scenario that will come to mind would involve the lowering of pain perception. Some sort of sham intervention would be administered, and extensive lies and surreptitious conditioning would combine to produce the desired placebo analgesia. Pain is, however, just one of the many dependant variables which placebo studies have sought to manipulate. Some of the manipulated variables are cognitive in nature; however, the number of studies on such outcomes is limited, so in this section I will cite all those I know of, to the benefit of the curious reader. My focus will be on effects produced on objective variables in healthy samples, but I shall mention the others as well.

Sustained attention placebo effects via verbal suggestion and false feedback. In two experiments conducted in 2009 by Colagiuri and Boakes, healthy participants were recruited for an alleged randomised placebo-controlled trial aimed at testing the cognitive-enhancing properties of a chemical compound. In fact, all received a placebo pill. Participants undertook a rapid visual information processing procedure, meant to measure sustained attention, three times: first as a baseline, then after taking the pill, and finally after having received false feedback on how their performance had either improved or remained the same after treatment. The third trial was thus the one of interest. Quite naturally, the participant would wonder whether they have been assigned the placebo arm or the intervention one; and the false feedback was meant to alter participants' beliefs regarding that question. Condition assignment belief was assessed at the end of the experiment. Those persuaded to have been assigned to the intervention arm outperformed the others, who believed they had been given a placebo, in terms of accuracy (first experiment) and reaction time (second experiment).

Placebo and nocebo effects on implicit learning via verbal suggestion and classical conditioning. Participants in a study by Turi and colleagues (2018) were asked to perform a reward-based learning task. They had to repeatedly choose between two symbols, one of which was associated with higher probability of reward than the other; through trial and error, participants would learn which symbol was better to bet on. Through a combination of verbal manipulation and classical conditioning, as well as through the employment of a sham non-invasive brain stimulation apparatus, the experimenters attempted to alter participants' expectancies, so that one group would believe the sham intervention would inhibit their performance, and another group, on the contrary, that it would improve it. The manipulation proved successful: participants' expectations about their own performance before the task varied according to group allocation, as did subjective impressions collected after task completion. Perhaps even more interestingly, the same pattern was found in actual accuracy; that is, participants manipulated to expect they would perform better or worse due to the (intrinsically inactive) intervention chose the most rewarding symbol more or less frequently, respectively. Turi and colleagues had already had similar findings a year before, when, using an analogous setup, they had produced an implicit learning placebo effect which had, likewise, improved both subjective and objective measures of performance (2017).

Inhibitory control placebo and nocebo effects via verbal suggestion and classical conditioning. Similarly, Schwarz and Büchel (2015) had probed whether a combination of verbal manipulation and classical conditioning might cause placebo and nocebo effects on inhibitory control. Their main finding, however, had been that, while subjective experience had been strongly affected, the experimental manipulation had not influenced objective measures of performance. An analogous pattern of results (impact on subjective measures, but not on objective ones) was later found by Winkler and Hermann (2019), who assessed the effect of a merely verbal manipulation on several cognitive dimensions (alertness, working memory, sustained attention and divided attention).

Motor learning placebo effect via verbal suggestion. A 2021 study conducted by Villa-Sánchez and colleagues saw participants perform a serial reaction time task. I.e., they had to press buttons as per visual instructions received, quickly and accurately; and, embedded among the random sequences of key pressings required, a deterministic sequence was hidden. The presence of these two components, random sequences and deterministic ones, allowed for the assessment of general motor learning and sequence-specific learning, respectively. Participants were divided among three conditions: one control group; one sham 'motor' intervention group, which had electrodes placed on the hand, and was explained that the light current would increase motor performance; one sham 'cognitive' intervention group, which had electrodes placed on the head instead, and was explained that the light current would increase attention and concentration. The two different placebo interventions affected performance differently: the 'motor' group displayed more accuracy and faster reaction times than the control one in the general motor learning part of the task, as well as lower subjective ratings of physical fatigue after the experimental manipulation; the 'cognitive' group, instead, registered less of both physical and mental fatigue after the manipulation, though no difference in actual performance. Fiorio and Emadi Andani, co-authors of the article, had already been involved together in two more studies on the topic of placebo effects on motor performance, both of which had found alterations on objective variables (Fiorio et al., 2014; Emadi Andani et al., 2015). Together, these results seem to rhyme with those concerning the existence of placebo and nocebo effects in sports, most recently demonstrated with a meta-analysis by Hurst and colleagues (2019).

Cognitive placebo effects in clinical populations. Two reviews have been written on placebo effects in psychiatric disorders in general (whose existence is beyond doubt: Weimer et al., 2015; Belcher et al., 2018). In addition, one meta-analysis focusing on schizophrenia was authored in 2017 by Keefe and colleagues, who registered a small but significant cognitive placebo effect whereby schizophrenic patients given placebo pills saw an improvement in their performance on the MATRICS Consensus Cognitive Battery (a measure designed to assess their speed of processing, attention/vigilance, working memory, verbal learning, visual learning, reasoning and problem-solving, social cognition: Nuechterlein et al., 2008). Another patient population in which cognitive placebo effects have been found is the one afflicted by Parkinson disease. In these patients, placebo and nocebo interventions were shown to modulate both positive and negative impacts of deep-brain stimulation (Keitel, Wojtecki et al., 2013; Keitel, Ferrea et al., 2013; Keitel et al., 2014), as well as reward learning (Schmidt et al., 2014).

A research plan

Peril

Search on Google Scholar for articles mentioning "placebo analgesia" in their title, and you will get about 508 results. That is for placebo analgesia alone, leaving aside all other placebo effects of medical interest. By contrast, my present count of studies on cognitive placebo effects – on *any* cognitive placebo effect – is 17 (including the accidental finding by Lowery and colleagues discussed above). It is, in short, a skewed process, that of formulating hypotheses concerning placebo effects, their psychological dynamics, their physiological correlates, or their moderators: whenever *general* theories of placebo effects are pondered, this is done with reference to a background knowledge - a literature – which has strongly privileged the exploration of one half of the wide realm of placebo effects *in general*, remaining ignorant of the other one. Such an unbalanced knowledge base puts us at risk of missing important hints about the nature of placebo effects, and of making hasty generalizations.

Dream

If solid evidence for one cognitive placebo effect could be found, and an easy-to-replicate experimental protocol to induce and measure it developed, that would not only be of interest in itself: it would mean to have opened for placebo research a whole testing ground for hypotheses and models first developed in the above-mentioned domains of medical relevance. And why not memory? Weber and Loughan's intuition that their finding may be worthy of the name 'placebo effect', however bold, was an inspiring one. A memory retrieval placebo effect would make for an ideal testing ground, in that it would be an unexplored one: because, while early evidence for these phenomena exists, their characteristics are still mostly unknown. Just by making small adjustments to the protocol, then, a cascade of purported mechanisms and moderators of placebo effects discovered in other domains could be tested outside of them, and thus what is domain-specific be distinguished from what features truly all placebo effects have in common… *aside from their denomination*.

Mission

To bring balance in placebo research, translate findings from medical domains to cognitive ones. To translate findings, create an effective and easily replicable protocol for the reliable generation of a cognitive placebo effect, a protocol of which variations could be flexibly produced to test a variety of theoretical models of placebo effects against a novel, unexplored phenomenon of the same family. A memory retrieval placebo effect would be ideal: such phenomena probably exist, but remain to this day obscure. Let the placebo impact episodic memory, rather than semantic one: it would take less time to run a protocol measuring retrieval in episodic memory (that which "receives and stores information about temporally dated episodes or events, and temporal-spatial relations among these events" (Tulving, 1972). Let the protocol be compatible with in-person and online research alike, for maximum chance of implementation. Share all study materials, obviously. How to elicit the placebo effect? Through which road pursue it? Two most classical options come to mind: either a manipulation of expectancy through verbal manipulation and false feedback – the most traditional choice, and the most expedient – or classical conditioning. In the end, it would be good to test both; let the first protocol employ classical conditioning, then, for it is the more technically challenging of the two. Therefore, the experiment shall comprise an association step, during which a stimulus will need to be systematically paired with easier retrieval. A pilot study will be needed, then: to assess the efficacy of the procedure meant to induce easier retrieval, first; and, second, provided that that manipulation check succeeds, to collect data on the size of the effect produced, so to attain an adequate estimate of the required sample size. This paper reports on that pilot study.

Ancestry

On the father side, this pilot study has an ancestor in Weber and Loughan's 2013 work, to whose conclusions it hopes to add. On the mother side, this pilot study descends from an already mentioned branch of experiments, that by Jensen and colleagues (2012, 2014, 2015), which featured the investigation of placebo effects via classical conditioning. Just like in those experiments, in this pilot study the conditioned stimulus had to be delivered subliminally, so to exclude expectancy as a potential placebo effect mechanism. Subliminal stimuli are most often delivered visually; however, when doing so, the stimulus duration must be set to be a multiple of the screen's refresh rate – which varies from one computer to another. The fact that the protocol had to be compatible with online research meant that many different computers would run it, and, therefore, visual stimuli seemed prone to compatibility issues. Thus, a 2005 experiment by Kouider and Dupoux, mentioned below, was called on to be this pilot study's preceptor and fencing teacher: it showed how to deliver speech stimuli subliminally.

Quest

At the most fundamental level, what this pilot study aimed for was the production of a procedure: whereby retrieval could be made easier outside of conscious awareness; whereby some retrieval trials would be made easier, and others not; whereby the trials made easier would be paired with a subliminal stimulus. An association would be formed in the mind of the participant – an instance of implicit memory (Graf and Schacter, 1985), between that stimulus and increased ease of retrieval. In a following step, that subliminal stimulus previously associated with easier retrieval would be paired with new trials. Ease of retrieval would be measured in these latter trials, through objective and subjective proxy measures alike, and compared with a baseline. If higher than baseline, then the pilot would have produced an evidence of retrieval placebo effects. Finally, in order to explore possible similarities between the phenomenon and others already established as placebo effects, as well as to better understand the phenomenon inner workings, two potential personality correlates would be measured: dispositional optimism, the most prominent candidate as moderator of placebo effects; and state anxiety, as Weber and Loughan hypothesised anxiolysis may have been an underlying mechanism of the memory enhancement effect they found. More precisely and exhaustively, this pilot study sought to test and answer these hypotheses and questions:

H1. Manipulation check: covertly suggesting the solution would improve performance.

H2. (Conditional on H1 being verified) an episodic memory retrieval placebo effect via classical conditioning would be registered.

H3. (Conditional on H1 being verified) an episodic memory subjective ease of retrieval placebo effect via classical conditioning would be registered.

H4. (Conditional on H2 being verified) a positive correlation would emerge between the observed episodic memory retrieval placebo effect and dispositional optimism.

H5. (Conditional on H3 being verified) a positive correlation would emerge between the observed episodic memory subjective ease of retrieval placebo effect and dispositional optimism.

H6. (Conditional on H2 being verified) a negative correlation would emerge between the observed episodic memory retrieval placebo effect and state anxiety.

H7. (Conditional on H3 being verified) a negative correlation would emerge between the observed episodic memory subjective ease of retrieval placebo effect and state anxiety.

Q1. Manipulation check: did participants remain unaware of the manipulation meant to produce increased ease of retrieval?

Q2. Manipulation check: did participants remain unaware of the conditioned stimulus?

Q3. Objectively and subjectively, how difficult was the task? Did it elicit either floor or ceiling effects?

Q4. Participants were expected to get tired as the experiment progressed. But when, and by which amount, specifically, would their fatigue increase? Would there be a gradual increase, or else a sudden one? When?

Q5. How did participants fare in the various areas of adherence to protocol? Was there some aspect that appeared particularly problematic?

Q6. Would the sound quality prove adequate? Would the participants be familiar with the words employed? Would they find the pronunciation adequately understandable?

Methods

Participants

A convenience sample of 21 volunteers was recruited via personal appeals among friends of friends and friends of relatives, as well as through two mailing lists including, respectively, students from Lund University and University College Cork. 11 participants self-identified as male, 10 as female. Mean age was 25.29 (SD = 7.25, range 20-59). In being invited to the study, and then again later, before consenting to participate, potential participants were informed that they would enter an experiment investigating the impact of music listening on speech processing. Also prior to consenting to participate, potential participants were made aware of the inclusion criteria: age above or equal to 18 years, normal or corrected-to-normal eyesight and hearing, belief that the surrounding environment would be quiet and free of distractions for the duration of the study (estimated in 85 minutes). All participants were debriefed at the end of the study and received my sincere thanks. The sample size was determined by availability.

Experiment design overview

Participants repeatedly performed a task that required them to memorise word pairs, first, and then to recall the second word of each pair, having been given the first as cue. The words were presented via audio recording, and in the recall step of the task many ‘nonwords’ (i.e., gibberish) would be played too, which served as masks for subliminal stimuli. One instance of that task (‘Baseline condition’) was used as baseline. In another instance of the task (‘Acquisition conditions’), half of the trials included the solution word as a prime embedded between masks, right before the cue word, to make retrieval of the solution word easier. In those trials there was also a specific nonword hidden between masks, which served as conditioned stimulus. The goal of including the conditioned stimulus was to create an association between the conditioned stimulus and increased ease of retrieval via classical conditioning. Following the Acquisition conditions, another instance of the task

(‘Conditioning condition’) took place, which included, hidden between masks, the conditioned stimulus, but not the solution words. I expected to find evidence of greater ease of retrieval in this Conditioning condition compared to Baseline condition. Had I found this, the result could have been interpreted as evidence of episodic memory retrieval placebo effect via classical conditioning.

The following sections describe, in increasing order of complexity: first, how the words and nonwords were selected and recorded; second, the exact structure of the memory task; and, third, the detailed anatomy of the entire experimental procedure. After that, a quick account of the open science practices adopted is offered, and a clear statement of the operationalised research hypotheses and questions.

Stimuli production and selection

The nonwords used in the experiment were created as follows. 2594 nonords (e.g., ‘ril’) were selected from the English Lexicon Project database (Balota et al., 2007) based on their length ("Length", 3-4 letters) and on how often they were recognised as nonwords ("NWI_Mean_Accuracy", 501-1000). They were pronounced and registered with an LS-10 Linear PMC Recorder (Olympus Imaging Europa BMGH, Germany) by the experimenter (Italian native speaker proficient in English). The silence in the recordings was truncated, the background noise reduced, the DC offset removed, the peak amplitude normalized to 0.0 dB (independently for each stereo channel). The experimenter then proceeded to further exclude 110 nonwords; 106 of them for being homophones to actual English words, the remaining four because of audio defects related to their registration. In this second selection process, the experimenter relied on the correct pronunciations provided by the Cambridge dictionary (Cambridge University Press, 2022). The resulting nonwords set thus consisted of 2,484 elements. Two copies of each nonword were at this point produced: a merely ‘compressed’ copy (whose beats per minute had been changed from 100 to 35, resulting in the same nonword being spoken in 35% of the original time, but without any change in pitch arising from the procedure, before its volume had been lowered by 15 dB) and a ‘reversed’ copy, which, having been accelerated and attenuated just like its 'compressed' twin, had also been reversed afterwards. All audio editing was accomplished with the free software Audacity (Audacity Team, 2021). Not all of those audio files were required in this experiment: out of the selected 2484, 1081 nonwords were randomly picked; 109 to be used in their compressed form, and 972 to be used in their 'reversed' one. The full list of all nonwords picked is specified in the shared experiment files (see 'Open science practices', below).

The selection and production of spoken word stimuli was more complex. Like for the nonwords, the written form of the word stimuli used was selected from the English Lexicon Project database (Balota et al, 2007). From a database of 79672 words, 798 were firstly selected for being nouns (POS=NN), common, singular, monosyllabic (NSyll=1), without homophones (Phono_N=Phono_N_H), of frequent use (SUBTLCD>=0.10). As a second selection step, three raters (two Chinese native speakers and the Italian experimenter, all proficient in English) independently read each of the words thus selected, noting, for each, whether they recognised it or not. Words recognised by at least two of the raters were selected for registration, the remaining were excluded; this resulted in the exclusion of 391 words. 407 words were thus registered to serve as stimuli. As it had been for the nonwords, it was the experimenter who spoke them (Italian proficient in English, male, 28 years old), using the same recorder, but this time making systematic use of the pronunciations made available on the Cambridge Dictionary (Cambridge University Press, 2022) as reference. As it had been for the nonwords, the silence in the recordings was truncated, the background noise reduced, the DC offset removed, the peak amplitude normalized to 0.0 dB (independently for each stereo channel) and the volume lowered by 15 dB. Employing the same procedure described above, 'compressed' copies of each word were made. Finally, the resulting 407 non-'compressed' audio files were divided into four sets of approximately equal numerosity (101-102), and each set was presented, in a random order, to four volunteers, whose task was to write down each word heard, after hearing it just once. While the volunteers were different for each set of 101-102 words, and no volunteer worked on more than one set, so to prevent fatigue effects, each set was submitted to one Italian native speaker and one Swedish native speaker who both had a self-rated

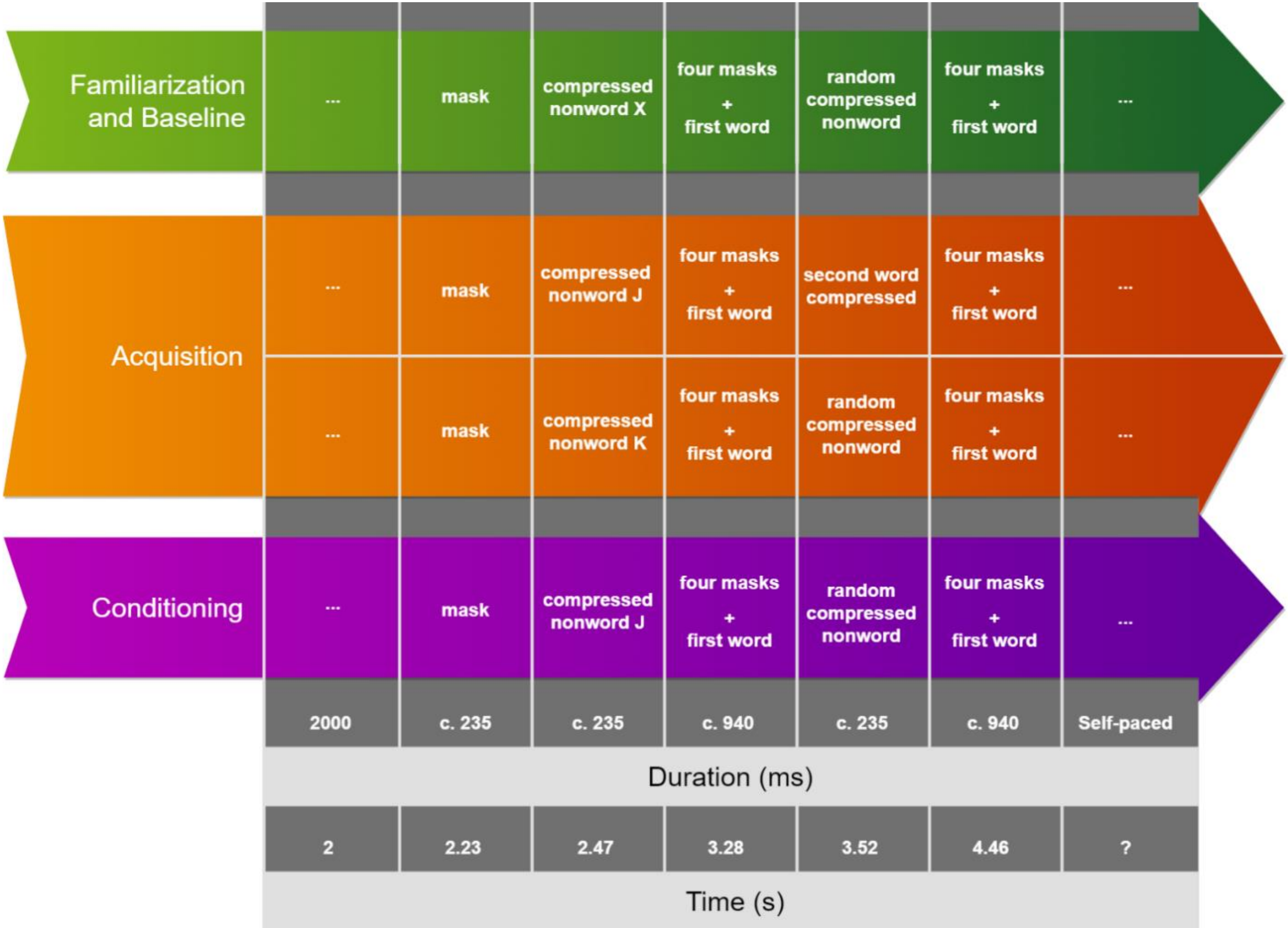
English CEFRL level of either B1 or B2 (Council of Europe, 2001), as well as to one Italian native speaker and one Swedish native speaker who both had a self-rated English level of either C1 or C2. 281 words were copied correctly by at least 3 of the 4 volunteers, and passed the selection. 17 words were copied correctly by only two volunteers, but at least one of the mistaken transcriptions appeared to the experimenter more like a mistake in how to write the word than a genuine confusion as to what the word meant; those were also accepted. Finally, the word 'mam', which had been transcribed as 'mum' four times out of four, also passed the selection. The remaining 108 registered words were excluded from use. Out of the 299 words selected, 216 were randomly picked, all of them to be used in both their ‘basic’ and ‘compressed’ forms. The full list of all words picked is specified in the shared experiment files (see ‘Open science practices’, below).

Musical Pairs

“Musical Pairs” is an adapted and rebranded version of the paired-associate learning task (e.g., Van Arsdall et al., 2014), and constituted the fundamental building block of the experiment. The task varied slightly depending on the specific conditions, but it always consisted of the same three steps. First, in Step 1 (encoding), participants were prompted to close their eyes and then exposed to 24 randomly formed word pairs to be remembered, in a random order Z (1, 2, 3, …, 24): for each pair, a bell sound was reproduced, then half a second of silence, then one ‘basic’ word, another half a second of silence, then another ‘basic’ word; a four-second encoding interval concluded each pair, during which participants could try to memorize the pair of words. Following the final word-pair came Step 2, which consisted of a one-minute-long delay period: smooth jazz music was played, an excerpt from “Bashful” (Ketsa, 2021), and visual instructions prompted participants to keep their eyes closed, relax, and focus on the music. The goal, with this step, was to erase working memory, as well as to reduce noise in data by making behaviour uniform between participants. Step 2 was ended by a question on whether any noise or distraction had caused the participant to lose a portion of Step 1. Third and last, Step 3 (retrieval) took place (Figure1). In this retrieval Step, participants first listened to the first word of a pair, as a recall cue; then, they were asked to remember the second word in the word pair. Subtle differences existed between Steps 3 in different conditions, but the general structure remained the same.

Figure 1

Step 3 (Retrieval): Outline of the Pattern of Auditory Stimuli, Varying Between Conditions


Familiarization and Baseline
...
mask
compressed nonword X
four masks + first word
random compressed nonword
four masks + first word
...
Acquisition
...
mask
compressed nonword J
four masks + first word
second word compressed
four masks + first word
...
...
mask
compressed nonword K
four masks + first word
random compressed nonword
four masks + first word
...
Conditioning
...
mask
compressed nonword J
four masks + first word
random compressed nonword
four masks + first word
...
2000
c. 235
c. 235
c. 940
c. 235
c. 940
Self-paced
Duration (ms)
2
2.23
2.47
3.28
3.52
4.46
?
Time (s)

Note. '…' indicates silence. By 'mask' is meant a random 'reversed' nonword (a different one at each instance). One different compressed nonword X was used for Familiarization and Baseline each; compressed nonwords J and K, instead, remained the same ones across all conditions (i.e., across all Musical Pairs rounds). 'First word' and 'second word' refer to the words in the word pair the participant had to memorise. Each Acquisition condition comprised 12 trials with compressed nonword J and 12 with compressed nonword K, in random order.

Like Step 1, Step 3 comprised 24 trials. Each of them would open with a silent visual reminder to close one's eyes and listen (a drawing of two closed eyes), lasting two seconds. After the eye closure, participants heard what appeared to them two repetitions of the first word (the recall cue) of one of the memorised word pairs, heard within a background noise of unintelligible babble. They had to respond to this by typing the second word of the memorized word pair. Unbeknown to them, the background noise was elaborately designed. In some conditions (called 'Familiarization' and 'Baseline' conditions) the background noise was truly unimportant and unhelpful for recalling the memorized word. However, in other conditions (called 'Acquisition 1' and 'Acquisition 2') there was a subliminal prime hidden in the babble in a random selection of 12 of the 24 trials, meant to make recall easier in these trials. Furthermore, in the same conditions, the presence of the subliminal prime was consistently paired with a specific nonword (conditioned stimulus), to achieve classical conditioning. Thus, this nonword became the conditioned stimulus for easier recall. In a fourth condition ('Conditioning condition') only the conditioned stimulus was hidden in the babble, without the subliminal prime, to measure the effectiveness of classical conditioning of easier recall. (The exact properties of each condition are explained in more detail below).

Here are, in order, the elements which made up the sound played to the participants after eye closure: a novel speech mask (one of the reversed nonwords described above); a speech prime (compressed nonwords described above); four more novel masks with the recall cue (first word of a pair) superimposed on them so that its beginning and that of the first mask would coincide; either a compressed nonword or the compressed version of the word which, in Step 1, had been second to the one now used as recall cue; and finally, like before, four novel masks with the recall cue superimposed on them. Based on Kouider and Dupoux's work, this procedure should have rendered the compressed nonwords, and the compressed second word, when present, subliminal. A complete audio example of an Acquisition Step 3, reproducing the pattern of sounds just described (including the hidden second word) can be downloaded and listened to from the web page https://github.com/Pietro-Rizzo/words_and_nonwords_1/blob/main/Acquisition_Step3_demo.mp3?raw=true.

After these audio stimuli, silence would follow, indicating participants it was time to open their eyes. Havin just heard the first word of one word pair, participants would then use their keyboard to enter the second word of that pair; or else to input 'no idea' or 'was disturbed', if that had been the case, although the instructions on the page would also prompt them to do so only as an extreme resort: if merely unsure, they should write their best guess. Having provided their answer, participants would move on to the next trial, revisiting all the 24 pairs of the previous Step 1. The order of the word pairs presented in the recall step would not be random, but follow an algorithm, designed to minimize and make constant item-order effects: if word pairs in Step 1 had randomly appeared for encoding in an order Z(1, 2, 3, …, 24), then here in Step 3 they would reappear for retrieval in a fixed order f(Z) (4, 6, 8, 1, 3, 5, 2, 7, 9, 11, 13, 10, 15, 12, 14, 16, 18, 23, 20, 22, 24, 17, 19, 21). Each single mask would be different, within and across trials; the nonword primes would instead remain constant within the Musical Pairs round (and possibly in other rounds as well, depending on the condition).

After the 12th trial, a measure of subjective difficulty was taken, as participants were asked to evaluate how much difficulty they were finding in this round, using a slider designed to imitate a visual analogue scale (VAS) not unlike those often employed in pain research, allowing 100 positions and displaying labels ranging from 0 to 10, with 0 being "none", 1 to 3 being "little", 4 to 6 being "medium" and 7 to 10 being "great". After the 24th trial, Step 3 would end with the Aftermath questionnaire, a series of eleven questions aimed at assessing to what extent the Musical Pairs round had proceeded according to plan (whether rules had been followed, whether any disruption had taken place, whether the participant had felt attentive, tired, etc. See Appendix A).

Procedure

The study consisted of eight sequences: Preliminaries (about 4 minutes), Familiarization condition (15 minutes), Baseline condition (13 minutes), Acquisition 1 condition (13 minutes), Acquisition 2 condition (13 minutes), Conditioning condition (13 minutes), Questionnaires (8 minutes), Debriefing (4 minutes). As the experiment design was within-subject, for counterbalancing purposes, the Baseline condition could take place either right after the Familiarization or the Conditioning one, with equal probability, at random. After every sequence but the first, the last, and the second to last, participants were invited to take a 5-minute break. The experiment was conducted online, in a single session lasting an average of 83,77 minutes (SD = 12,66 range 64-112), with English as the sole language employed. The open-source software Open-Sasame (Mathôt et al., 2012) was used to program and run the experiment. The experiment was deployed online through JATOS (Lange, Kühn & Filevich, 2015) on the free server for hosting online experiments MindProbe (courtesy of the European Society for Cognitive Psychology and OpenSesame).

Figure 2

Experiment outline

Procedure (Baseline first)	Procedure (Baseline last)	Sequence duration (minutes)	Total time (minutes)
Preliminaries	Preliminaries	4	4
Familiarization	Familiarization	15	19
Baseline	Acquisition 1	13	32
Acquisition 1	Acquisition 2	13	45
Acquisition 2	Conditioning	13	58
Conditioning	Baseline	13	71
Questionnaires	Questionnaires	8	79
Debriefing	Debriefing	4	83

Preliminaries

Having opened the study link, participants were invited to set their browsers to fullscreen mode, note down my email address for future questions or bug reports, wear headphones if available and turn off their phones. Participants then confirmed their conformity to the inclusion criteria and gave explicit consent to participate. A short audio volume calibration sequence followed, at the end of which participants were instructed not to alter the audio setting until the end of the study.

Familiarization

After that, participants were described the three steps of Musical Pairs, the “game” “a few rounds” of which they would be required to play through the experiment:

1) first, you will close your eyes and hear several word pairs (for example, you may hear 'mirror... dog');
2) second, you will listen to some music;
3) third, the first word of each pair will be played twice, and you will have to write down the corresponding second one (for example, you would hear 'mirror... mirror', and you would need to type 'dog').

Furthermore, two additional instructions were provided and repeated. First, during each Step 1, while listening to word pairs (i.e., encoding) participants were to ‘try to see in their mind whatever image each word pair inspires them’; this was exemplified twice, for instance with the word pair ‘fig… boat’, which participants were suggested could be visualized as “a boat full of figs”. Second, participants were warned that during each Step 3 (i.e., retrieval) there would be “other noises” played in the background, and not to be distracted by it; they were explained that those were

> distorted nonwords; that is, word-like speech sounds that are completely meaningless. Like 'atup', or 'blim'. You do not have to worry about them. I put nonwords there to make the experiment more like what it would be to listen to someone in a crowded area in the real world, such as a cafeteria, where there's always some background chatter. Our brains process words in a particular way when they are heard close to other speech-like sounds, and it is that particular way of neural processing that I am interested into.

As far as the Familiarization condition was concerned, the noises were indeed nonwords. See Table 1. Nine always novel masks, one Familiarization-specific recurring ‘compressed’ nonword (prime) X, one always novel compressed nonword.

The Familiarization condition included three short practice rounds, with pieces of instructions being presented and repeated as needed in the first round, and less and less guidance being provided as the participant progressed to the second and then third round, acquiring experience.

Baseline

The Baseline condition consisted of a standard Musical Pairs round. The characterizing features were the same as those of the shorter Familiarization round: for each trial, nine novel masks, the same Baseline-specific recurring ‘compressed’ nonword (prime) X and one novel compressed nonword. See Table 1.

Acquisition 1 and 2

The condition Acquisition 1 differed from Baseline in that instead of one Baseline-specific prime recurring at every trial, one of two compressed nonwords J and K were played each time. Following that, if K had been played, then a random compressed nonword would be subliminally played as well, just as during Baseline. But if J had been played, then the second word of the trial's pair would be played instead, in 'compressed' form, to subliminally prime recall of that word. 12 pairs were presented together with each nonword, in random order, for a total of 24 pairs. This procedure was meant to produce an association between nonword J and increased ease of retrieval without eliciting the formation of expectancies (i.e., outside of conscious awareness). Whether this would work was the first of the main questions this experiment was hoped to answer. Acquisition 2 followed the same pattern as Acquisition 1, and employed as primes the same two 'compressed' nonwords K and J.

Conditioning

The Conditioning condition was modelled after the Baseline one, except that the prime in each trial was that same 'compressed' nonword J that had been associated with subliminal help (i.e., the second word of the pair) being provided during Acquisition 1 and 2. This time, however, no help was given. This condition's aim was to measure the effectiveness of the classical conditioning of easier recall which was achieved in the Acquisition conditions by consistently pairing the subliminal priming of recall with the conditioned stimulus (nonword J).

Questionnaires

At the conclusion of the last condition (either Conditioning or Baseline, depending on the counterbalancing random variable), participants completed a few questionnaires.

Life Orientation Test-Revised (LOT-R). The first questionnaire to be completed was the Life Orientation Test-Revised (LOT-R; Scheier et al., 1994), a short measure investigating generic positive expectancies and assessing dispositional optimism. The LOT-R consists of 10 self-report items, each rated on a 5-point scale ranging from zero (strongly disagree) to four (strongly agree): three expressing a pessimistic outlook on life (e.g.: “I hardly ever expect things to go my way”), three expressing an optimistic outlook on life (e.g.: “I’m always optimistic about my future”) and four neutral filler items. The test is scored by reverse scoring the three pessimistic items and averaging them together with the three optimistic ones. The measure has consistently shown adequate psychometric properties, with Cronbach’s alphas ranging between .66 and .80 (Hinz et al., 2017, Chiesi et al., 2013; Glaesmer et al., 2011).

Brief State-Trait Anxiety Inventory – State (BSAM). State anxiety was then assessed through the State scale of the Brief State-Trait Anxiety Inventory (BSAM; Berg, Shapiro, Chambless & Ahrens, 1998). This is a shorter version of the second scale of the State-Trait Anxiety Inventory, form Y (STAI-II; Spielberg et al., 1983), to which it has been shown to highly correlate (r = .93), while demonstrating good internal consistency (alpha = .83 and .86). The scale consists of 6 self-report items concerning the participant’s present state (e.g., “I feel relaxed”, “I feel steady”), each rated on a 4-point scale: 1 (“almost never”), 2 (“sometimes”), 3 (“often”), 4 (“almost always”). After the three positive items are reversed, a state anxiety score ranging from 6 to 24 is computed, with higher numbers indicating higher state anxiety.

Musical Pairs Questionnaire. An ad-hoc questionnaire followed, meant to investigate audio quality through the Musical Pairs tasks, as well as pronunciation understandability and words familiarity. The Musical Pairs Questionnaire is presented in Appendix B.

Suspicion Questionnaire. An ad-hoc series of three questions followed, probing for suspicion. Participants were first asked whether they had noticed anything different between the four longer Musical Pairs rounds (open question); then, having been reminded of the many short, distorted, speech-like sound they had heard, participants were asked whether they had sometimes noticed anything unexpected about them, and requested to be as specific as possible (open question). Finally, being more specific, they were asked whether they had noticed that one of the nonwords in each trial was always a number, to which they could reply either yes or no. This last question was meant as a trap to identify untrustworthy respondents.

English Questionnaire. The next set of questions revolved around the participant's language skills. The first of three questions inquired about the participant's native language(s), the second about the extent of their English command, the third, for English native speakers, about their accent.

Demographics. Finally, three quick questions on demographics prompted the participant to state gender, age, and highest level of formal education attained.

Debriefing

Participants were then fully debriefed, thanked, and given the opportunity to enter a separate survey in which to indicate an email address for them to receive a link to the paper describing the results of the study they had contributed to, once that paper would have been ready. As this was accomplished through a separate and completely anonymous survey, the emails thus collected were in no way connected with the rest of the data provided.

Open science practices

Due to its explorative and tentative nature, the study was not pre-registered. The song excerpt used in Step 2, the experiment code and the 5782 audio files of words and nonwords the experimenter produced for stimuli are freely available to everyone at https://github.com/Pietro-Rizzo/words_and_nonwords_1 (Rizzo, 2021). The experiment can be experienced at https://jatos.mindprobe.eu/publix/3506/start?batchId=3987&personalMultipleWorkerId=75381.

Research hypotheses and questions

The hypotheses and research questions on which this pilot study was based (expressed above), where operationalised as follows:

H1. Within participants, the number of recalled words will be higher in Acquisition trials including the subliminal prime compared to Acquisition trials not including the subliminal prime. (Participants who report consciously hearing the subliminal prime are excluded from this analysis). Tested with a one-tailed paired sample t-test.

H2. (Conditional on H1 being verified). Within participants, number of recalled words would be lower in the Baseline condition than in the Conditioning condition. (Participants who report consciously hearing the subliminal prime or being aware of the conditioned stimulus are excluded from this analysis). Tested with a one-tailed paired sample t-test.

H3. (Conditional on H1 being verified). Within participants, subjective difficulty scores would be lower in the Baseline condition than in the Conditioning one. (Participants who report consciously hearing the subliminal prime or being aware of the conditioned stimulus are excluded from this analysis). Tested with a one-tailed paired sample t-test.

H4. (Conditional on H2 being verified). The difference of number of recalled words in the Conditioning vs. Baseline condition will be positively correlated with dispositional optimism. Tested with correlation analysis (Pearson's *r*).

H5. (Conditional on H3 being verified). The difference of subjective difficulty score in the Conditioning vs. Baseline condition will be negatively correlated with dispositional optimism. Tested with correlation analysis (Pearson's *r*).

H6. (Conditional on H2 being verified). The difference of number of recalled words in the Conditioning vs. Baseline condition will be negatively correlated with state anxiety. Tested with correlation analysis (Pearson's *r*).

H7. (Conditional on H3 being verified). The difference of subjective difficulty score in the Conditioning vs. Baseline condition will be positively correlated with state anxiety. Tested with correlation analysis (Pearson's *r*).

Q1. How many participants, out of the total number, would report hearing the subliminal prime (i.e., report the solution word being suggested in some trials)?

Q2. How many participants, out of the total number, would notice that some nonwords had been repeated?

Q3. What were the mean and standard deviation of participants' total recall score? What were the mean and standard deviation of participants' subjective difficulty score?

Q4. Can statistically significant differences be observed between participants' ratings of fatigue across the first, second, third and fourth condition (i.e., through time, rather than depending on the properties of conditions)? Tested with one-tailed paired sample t-tests.

Q5. What were the means and standard deviations of participants' answers to each of the items of the Aftermath questionnaire?

Q6. What were the means and standard deviations of participants' answers to Musical Pairs questionnaire items?

Results

Technical issues

While the software used in the present experiment had been programmed to play all sounds one after another, without any delay between them, the actual output always included brief moments of silence between sounds. Measuring it exactly was not possible, but that gap appeared to last about 600 ms.

Furthermore, five people reached out to me to document technical issues which had prevented them from completing the experiment. They would have been, otherwise, five more participants, in addition to the aforementioned 21. Among their reports, two were from participants who could not proceed beyond the sound test, while other two concerned audio temporarily ceasing to work, only to then resume its functioning and reproduce several ‘overdue’ sounds either at the same time or at an accelerated speed. The fifth came without details.

Q1, Q2. Manipulation checks: unawareness of the two manipulations

10 out of 21 participants correctly reported that in some trials the solution was being suggested between one mask and another. No participant, however, reported on the compressed nonword primes which had being repeated through multiple trials. Only one participant reported having noticed that a number had always been played through each different trial, and that participant had, just to the previous item, answered negatively to the question as to whether there had been anything unusual about the nonword sounds in the background.

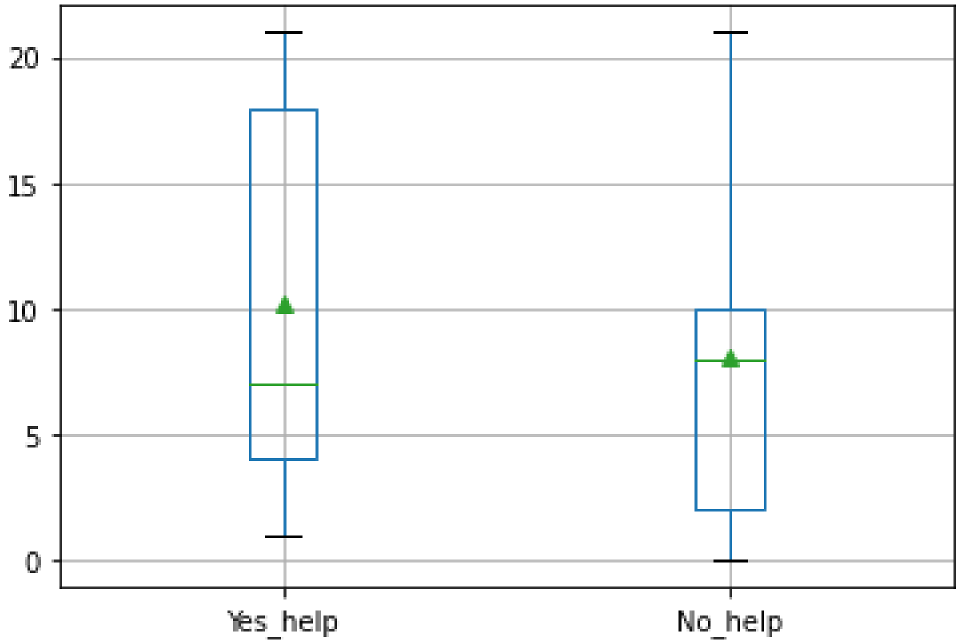
H1. Manipulation check: covertly suggesting the solution would improve performance

Among those 11 who had not reported being aware of the subliminal prime (i.e., the solution word being suggested), the average score for all trials with the subliminal prime was 9.54 (SD = 7.83); while the score for trials without the subliminal prime was inferior: 8.27 (SD = 6.59); this difference, however, was not significant (t = -1.58, df = 8, p = .077; paired t-test, one-tailed). See Figure 3. (A paired, one-tailed t-test considering instead all 21 participants (t = 1.78, df = 20, p = .045) showed that the (not-much-)subliminal prime had the intended significant positive effect on recall (primed trials: M = 10.90, SD = 7.02; non-primed trials: M = 8.86, SD = 6.48)).

Participants who were aware of the subliminal prime had a higher mean score for the trials with subliminal prime (M = 12.40, SD = 6.06) than those who did not report being aware of the subliminal prime (M = 9.54, SD = 7.83), but the difference was not significant (t = .94, df = 18, p = .179; Welch t-test, one-tailed). Likewise, participants who reported awareness of the subliminal prime had a higher mean overall score than the others across all four Musical Pairs, the respective means being 45.7 (SD = 22.45) and 33.91 (SD = 25.40); the difference was not significant (t = 1.13, df = 18, p = .137; Welch t-test, one-tailed).

Figure 3

Performance of naïve participants across trials with and without help



Note. Performance of participants who did not report awareness of hidden help across J-prime trials (where help was provided, i.e., the second word was suggested) and K-prime trials (where help was not provided). Triangles indicate the mean.

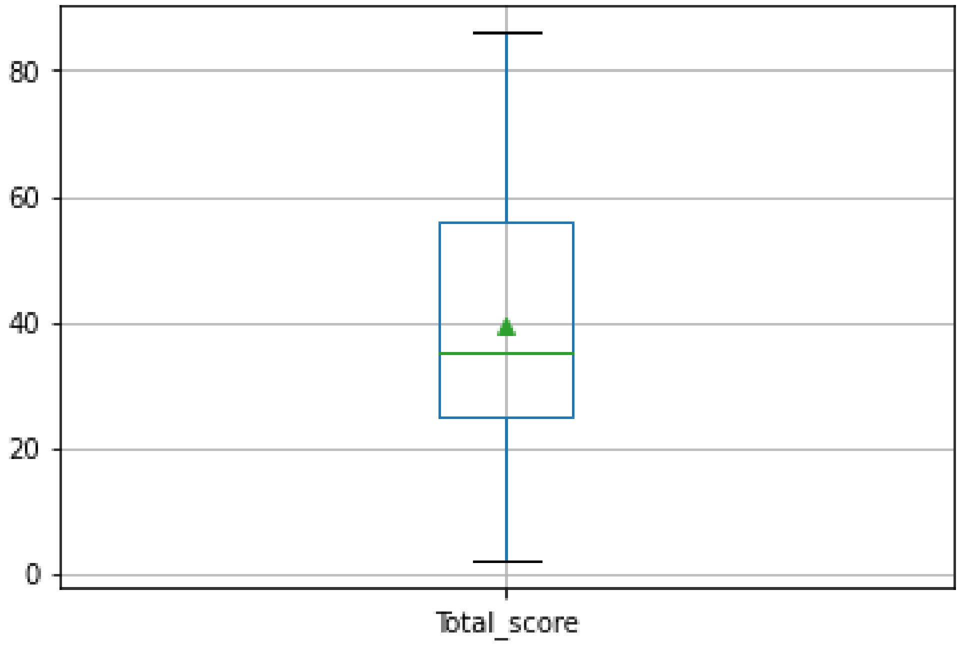
Q3. Task difficulty

Mean subjective difficulty ratings (1-100) across conditions were as follows: 60.52 for Familiarization (SD = 19.22), 62.57 for Baseline (SD = 15.71), 66.57 for Acquisition 1 (SD = 15.51), 63.57 for Acquisition 2 (SD = 14.80), 66.19 for Conditioning (SD = 18.12). No significant differences between conditions were registered: $F(100,4) = .84$, $p = .506$.

Participants' objective performance was measured as the number of correct guesses of the second word of the pair tested; any answer differing from the exact spelling of the solution word was treated as incorrect. On a qualitative note, this meant excluding and grouping together answers that appeared to the experimenter as: words that sounded similar to the solution word ("couch", "pouch"); words semantically related to the solution word ("corpse", "dead"); words that sounded similar to the solution word which were also semantically related to it (“grid”, “grate”; “player”, “play”). Individual performance varied greatly between participants, with the grand total of correct answers across all four Musical Pairs ranging from 2 to 86 out of theoretical maximum of 96, the mean being 39,52 (SD = 24,21). See Figure 4.

Figure 4

Overall performance of all participants.



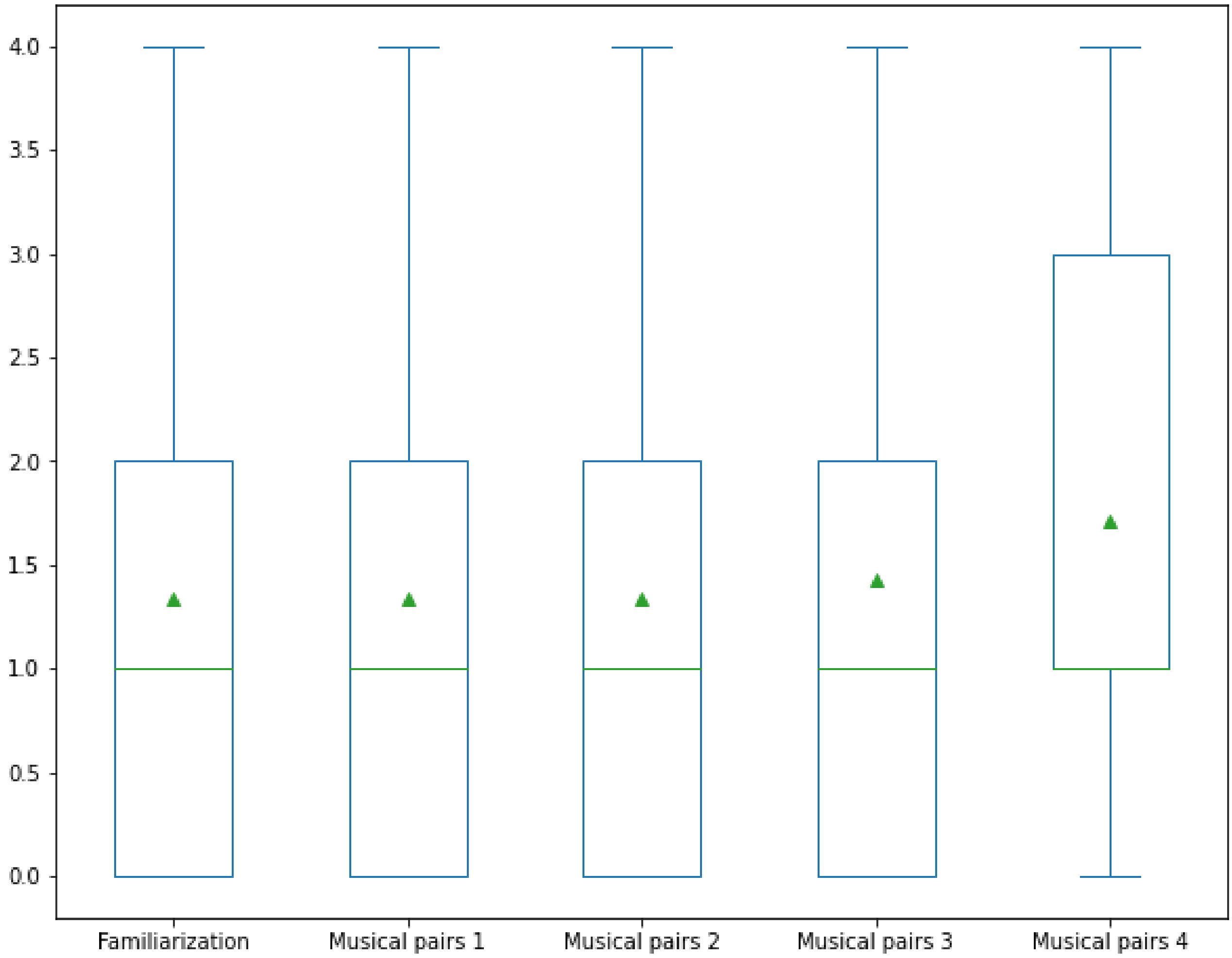
Note. Overall performance of all participants. The triangle indicates the mean.

Q4. Fatigue through time

Participants' ratings of perceived tiredness ("I feel tired", Item 6 of the Aftermath questionnaire, with values ranging from 0 to 4) display an ascending trend (Fig. 3). Specifically, ratings collected after the fourth Musical Pairs round (i.e., the last: either Baseline or Conditioning) significantly differ from each of those collected after each other Musical Pairs task: the first (t = -2.17, df = 20, p = .021), the second (t = -2.02, df = 20, p = .029), and the third (t = 1.83, df = 20, p = .041; all t-tests being paired and one-tailed). No other difference was significant. However, a Welch t-test showed that no significant difference in Baseline performance existed between those participants who had taken the measurement before the other four Musical Pairs and those who had taken it afterwards.

Figure 5

Reported feeling of tiredness through time



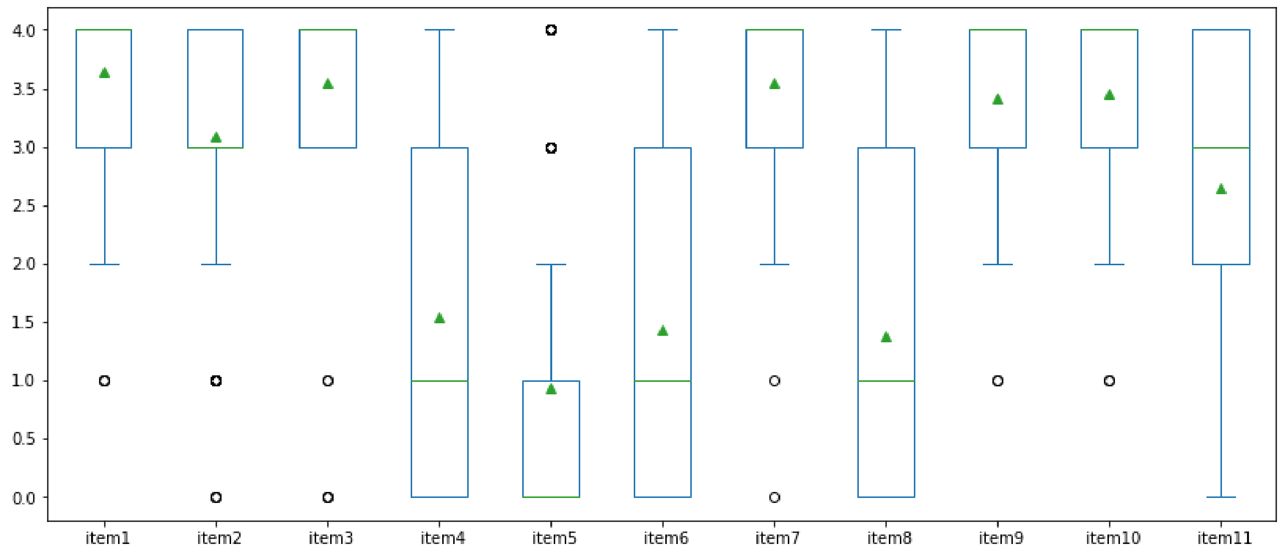
Note. Reported feeling of tiredness through time, on a scale from 0 to 4. Triangles indicate the mean.

Q5. Participant adherence to protocol

Participant' self-reported degree of agreement with each statement of the Aftermath questionnaire are described in Figure 6. For items 4, 5, 6 and 8, denoting qualities of a poor experiment session, all means across conditions were lower than 2. For the remaining items, denoting qualities of a good experimental session, no mean value was lower than 3, with the sole exception of Item 11 ("During Musical Pairs, as the music played, I just listened, thinking of nothing in particular"), whose mean was 2.64 (SD = 1.45).

Figure 6

Answers to the aftermath questionnaire, irrespective of condition



Note. Triangles indicate the mean. For the content of items, see Appendix A. Note how items 4, 5, 6 and 8, which are the only ones denoting undesirable qualities in participation (unrelated thoughts, interruptions, tiredness, noisy environment), are also the only ones whose average scores converge near the bottom, while all other items, denoting desirable qualities, display ceiling effects instead. Also note how this is less true for Item 11 (focus on music alone during Step 2).

Q6. Language of participants

With the Musical Pairs Questionnaire (see Appendix B), participants were given an opportunity to highlight some barriers to word comprehension they experienced during the experiment. The results are presented in Table 1. Items 1 to 3, concerning audio quality, elicited mean responses either above 3 or below 2. Items 4 and 5 concerned participants' familiarity with the words used; the mean answer to item 4 is above 3, but that to item 5 is close to neutrality. The same can be said for items 6 to 8, all concerning pronunciation.

Table 1

Summary of the answers to the Musical Pairs Questionnaire

Item	Content summary	M	SD
Item 1	Sound was crisp.	3,24	0,77
Item 2	Words were loud enough.	3,38	0,67
Item 3	Sound was distorted or metallic.	1,38	1,24
Item 4	I was familiar with the words.	3,24	0,70
Item 5	Some words were new to me.	2,71	1,10
Item 6	Words were pronounced clearly.	2,86	1,01
Item 7	There was a heavy accent.	2,62	1,40
Item 8	Difficult to understand due to pronunciation.	2,57	1,33

11 participants identified as native English speakers (10 having an Irish accent, 1 having a North American one); four as Italian, three as Swedish, two as German. Converting CEFRL levels into numbers, with A1 = 1 and C2 = 6, the mean self-reported language level was 4.44 (SD = 1.42).

H2 and H3. Evidence of semantic memory retrieval placebo effect

H2 could not be tested, because it was conditional on a significant effect on H1. Out of those 11 who had not reported the second word being suggested, three participants scored lower, one scored the same, and seven scored higher when receiving it (J-prime trials) than when not receiving it (K-prime trials) during Acquisition. When the Baseline scores of that latter subset of seven were compared to their respective Conditioning scores, no significant difference emerged ($t = .60$, $df = 5$, $p = .714$; paired t-test, one-tailed).

H3 could not be tested, because it was conditional on a significant effect on H1. Out of those 11 participants who had not reported the second word being suggested, five judged the Baseline condition to be more difficult than the Conditioning one; five other felt the opposite, and for one participant the two conditions were equally challenging.

H4, H5, H6, H7. Optimism and state anxiety

Hypotheses 4-7 could not be tested, because it was conditional on a significant effect on H2 and 3. Optimism mean score was 11 (SD = 5.08). State anxiety mean score was 12.09 (SD = 3.71).

Discussion

Present findings and future attempts

This pilot study was aimed, first and foremost, at investigating whether episodic memory retrieval could be facilitated through subliminal speech primes. Furthermore, it sought to assess whether that facilitation could be employed to elicit an episodic memory retrieval placebo effect via classical conditioning, and to measure the size of that placebo effect, as a basis for future sample size calculations. This pilot study was also meant as a way to test the feasibility of an experimental protocol quite rich of novel, untested components. The main actual usefulness of this pilot study, in the end, turned out to be on that latter side, of probing feasibility. Its results inspire several ideas on how a similar future study might be developed in order to effectively probe the research questions its predecessor could not.

First, the composite set of audio stimuli in Step 3 should have been subjectively experienced by participants as nothing more complex than two clear repetitions of the recall cue, “surrounded by unintelligible babble” (i.e., a noise “not unlike that of a background conversation”; Kouider and Dupoux, 2005, p. 617). However, a somewhat erratic gap of about 600 ms appeared between sounds.

It appears very likely that such undesired gap might have led to the likewise undesired awareness of words meant to be subliminal experienced by at least 10 out of 21 participants. That of temporal precision in stimuli delivery and in data collection, and of its insufficiency, is a point presently discussed among those who study and advance the state-of-the-art in online psychology research (Anwyl-Irvine et al., 2020). Most unfortunately, it appears that the longest average lags are produced when audio stimuli are employed (Bridges et al., 2020). Any future attempt at improving on this study should aim at preventing all such lags. One way to accomplish that could be to merge all the masks, nonwords and words used in each individual Step 3 trial (a total of 12 sounds) into a single audio file. This would translate into a trade-off between that randomization of stimuli which is meant to protect the experiment design from item-order effects, on the one hand, and effectiveness of manipulation, on the other. The loss in randomization could be compensated by having different versions of the protocol, differing in sounds order only, to which participants could be randomly assigned.

Second, technical issues have emerged beyond the above. The recruitment strategy and the data collected do not allow for a clear assessment of the number of potential participants approached, nor of the number of participants who dropped out from the study in general, nor of the number of participants who dropped out from the study due to technical issues; a lower bound on that last number may be established, however, by tallying the issue-reporting emails sent to the experimenter, which is five. Issues concerning sound such as those reported by four people seem problematic as well as difficult to solve, as no cause has been identified yet. If it is, on the one hand, lucky that they would appear to prevent data collection, rather than confounding the results, on the other hand the loss of participants is clearly undesirable, and the wasting of the participant’s time even more so.

Third: moving from software malfunctioning to human behaviour, it is noted that participants often appeared to misunderstand which words were spoken, and consequently provide incorrect answers. Results from the Musical Pairs Questionnaire tell us that the audio quality is not an issue. They can also be interpreted as evidence that the average participant was familiar with the vast majority of words, though being ignorant of a few. But, above all, they seem to suggest that a discrepancy may exist between the experimenter's pronunciation and that of the participant, a discrepancy likely to make the task much more difficult. Given the effort expended into producing understandable audio files, this is probably better explained by differences in accent, rather than objectively poor pronunciation or audio quality. The problem may be solved in data analysis by having a panel of three trained judges establish whether each such answer should be treated as an instance of correct remembering of a misunderstanding; but far better (for speed and reproducibility both) it would be for those misunderstandings not to present themselves in the first place. A future study may accomplish that by having participants keep their eyes open and displaying the written words as they are spoken during Step 1 (encoding), so to leave no room for ambiguity. This is particularly desirable in that judges marking a word that sounds similar to the solution as correct would generate a confusion between the remembering of sounds and the remembering of meaning. Participants may hear 'fraud' correctly during encoding, then make a mistake and write 'road' instead. Or, they might understand 'road' from the start and correctly remember it. As an alternative, one could either have all the words re-recorded by a member of the population from which the sample will be drawn, or else try to find a sample whose English accent is similar to the one in which the already registered words are spoken.

Fourth, there were no exclusion criteria to this study. They were not set because of the explorative and tentative nature of this first attempt, so to leave room for the unfiltered variety of the participants' population, which was yet to be understood. Once all issues will have been corrected, however, a confirmatory study will be more desirable, and that will require exclusion criteria. Several options come to mind. First, a lower bound to correct answers may be set, so that participants scoring lower than that are excluded form data analysis; for, what could cause a participant to score two out of 96 may forever remain a mystery, but we can guess it is something likely to interfere with the phenomena being studied. Whatever the arbitrary cut-off, it should refer to each individual Musical Pairs round, rather than the total, so that data from participants whose performance falls below the threshold even for just one of the four trials are in their entirety excluded from analysis. For the same reason, another exclusion criteria should be to remove from analysis all participants who ever report disrupted encoding, as this has the potential to act as a confound. Finally, two further exclusion criteria could be rooted in answers to the three questions probing for suspicion: participants who notice the sounds supposed to remain subliminal will have to be excluded, and it may be wise to also exclude those who provide a positive answer to the question as to whether they had noticed that all trials contained a number hidden among nonwords. (Although, based on the data collected in this occasion, that last criterion would be unlikely to be used too often; moreover, to avoid misclicks, that question about numbers would need to be changed to allow for an open answer, as the other two probing for suspicion).

Fifth. Being an online experiment, the study relied more heavily than usual on participants' self-reports. This is an unavoidable weakness, unless one wishes to abandon the possibility of online research. From those self-reported variables, the picture emerges of study conditions (environmental quiet, focus, etc.) that are overall well-suited to data collection; if there is one finding that less than others conforms to that satisfactory picture, it is that of participant behaviour during Step 2, i.e., the delay task (listening to music). That may be due to the excessively complex phrasing of the item meant to assess the behaviour ("During Musical Pairs, as the music played, I just listened, thinking of nothing in particular"), which would in any case be a good idea to simplify and make less ambiguous before future uses. Nonetheless, the observed relatively low scores may reflect a problematic reality. The desired behaviour was for participants to focus on music, and the rationale for that was that uniformity of mental activity and behaviour during the delay period would prevent noise from being added to the data. Not providing something to do during that short break might have resulted in participants engaging in all sorts of different activities; in the worse-case scenario, some would have rehearsed the word pairs, while some others would have not. Following such a line of reasoning, it may prove useful, when planning a variation on this study, to consider substituting the listening to the music task with a more interactive one, such as reading a story or, to better keep distance from all the meanings encoded, perform arithmetic.

There is more to learn from this experiment. For instance, this study succeeded in tracking the emergence of feelings of tiredness, which arose after the fourth Musical Pairs round: between 58 and 71 minutes into the experiment, then, it could be estimated. While this did not appear to impact performance, the information is to be kept in mind for future use, as a recommended upper bound for the duration of a similar study.

Furthermore, once the ambiguity in word pairs encoding is corrected (as described above) it may prove interesting to code answers in a more varied fashion than simply right or wrong. As the task presents chances for various kinds of mistakes, it might be worth exploring and quantifying the qualities of mistakes and of those who make them, and in which conditions.

Moreover, the study's main mechanic (Musical Pairs) produced an ample and well-distributed variety of measurements in the main dependant variable (number of correct recalls). This clear absence of either floor or ceiling effects is very much desirable, as it would allow for the measurements of manipulation-induced variations in the distribution. At the same time, based on the surprising amount of variance between participants, it seems recommendable that future studies follow in the tracks of this one, if feasible, by adopting a within-subject design.

A future study could, as envisaged above, attempt to produce a placebo effect on episodic memory retrieval via manipulation of expectancy. Such a study may follow the example of Weber and Loughan (2013) and seek to persuade participants that the word to be remembered is subconsciously available to them. That study may also measure and control for reliance on gut feelings, so to disentangle placebo effect dynamics from emotional oracle ones (see Pham et al., 2012, discussed above); and perhaps assess participants' suggestibility with questions such as 'How many times did you feel you were guessing?'.

Finally, one notes that, among the 11 participants who did not report noticing the help word, the difference between J-prime trials (with help) and K-prime trials (without help) was, if not significant, at least in the expected direction. This does not really mean much, of course, but it is somewhat encouraging for the eventual feasibility of producing a facilitation of retrieval through subliminal speech.

Conclusions

This ‘episodic memory placebo’ line of inquiry has always been, since the first study design, quite the plunge into the dark. Not only the experimental protocol is – was, now – largely untested, but the study hypothesis itself is thrice removed from its anchoring to the literature. Once, because in the literature both placebo analgesia via subliminal conditioning (Jensen et al. 2012, 2014, 2015) and ‘academic placebo’ effects (Weber and Loughan, 2013; Lowery, 2007) operate a manipulation through visual stimuli, not auditory ones; twice, because what is documented in the literature is, at best, a semantic memory placebo, not an episodic memory one; and thrice, because even that is elicited through a manipulation of expectancy, not classical conditioning. In all fairness, the anchoring itself cannot exactly be called strong: two studies, providing, together, first evidence of a related phenomenon. Is that something Karl Popper (1962, p. 227) would praise as a ‘bold conjecture’? Hopefully.

This study accomplished nothing in the direct advancement of that ‘episodic memory placebo’ line of inquiry; it has, however, been its starting point. Through this work, critical issues in the experimental protocol have been brought to light, and unforeseen new spaces for improvement sighted for the first time. Whatever the unknown destination of this research path, this experiment will have been a necessary step. Onward!

References

Adamczyk, W. M., Wiercioch-Kuzianik, K., Bajcar, E. A., & Bąbel, P. (2019). Rewarded placebo analgesia: A new mechanism of placebo effects based on operant conditioning. *European Journal of Pain*, *23*(5), 923–935. https://doi.org/10.1002/ejp.1360

Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods*, *53*(4), 1407–1425. https://doi.org/10.3758/s13428-020-01501-5

Audacity Team. (2021). *Audacity(R): Free audio editor and recorder* (Version 3.3.0) [Computer software]. https://www.audacityteam.org

Bąbel, P. (2019). Classical conditioning as a distinct mechanism of placebo effects. *Frontiers in Psychiatry*, *10*. https://doi.org/10.3389/fpsyt.2019.00449

Bajcar, E. A., & Bąbel, P. (2018). How does observational learning produce placebo effects? A model integrating research findings. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.02041

Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*, 445–459.

Belcher, A. M., Ferré, S., Martinez, P. E., & Colloca, L. (2018). Role of placebo effects in pain and neuropsychiatric disorders. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *87*, 298–306. https://doi.org/10.1016/j.pnpbp.2017.06.003

Benedetti, F. (2014). Placebo Effects: From the Neurobiological Paradigm to Translational Implications. *Neuron*, *84*(3), 623–637. https://doi.org/10.1016/j.neuron.2014.10.023

Benson, M.D, H., & Friedman, Ph.D, R. (1996). Harnessing the power of the placebo effect and renaming it “remembered wellness”. *Annual Review of Medicine*, *47*(1), 193–199. https://doi.org/10.1146/annurev.med.47.1.193

Berg, C. Z., Shapiro, N., Chambless, D. L., & Ahrens, A. H. (1998). Brief state-trait anxiety inventory-State [Database record]. Retrieved from PsycTESTS. doi: https://dx.doi.org/10.1037/t14279-000

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*, e9414. https://doi.org/10.7717/peerj.9414

Cambridge University Press. (2022). *Cambridge English dictionary: Meanings & definitions*. https://dictionary.cambridge.org/dictionary/english/

Chiesi, F., Galli, S., Primi, C., Innocenti Borgi, P., & Bonacchi, A. (2013). The Accuracy of the Life Orientation Test–Revised (LOT–R) in Measuring Dispositional Optimism: Evidence From Item Response Theory Analyses. *Journal of Personality Assessment*, *95*(5), 523–529. https://doi.org/10.1080/00223891.2013.781029

Colagiuri, B., & Boakes, R. A. (2009). Perceived treatment, feedback, and placebo effects in double-blind RCTs: an experimental analysis. *Psychopharmacology*, *208*(3), 433–441. https://doi.org/10.1007/s00213-009-1743-9

Colloca, L., & Miller, F. G. (2011). How placebo responses are formed: a learning perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1572), 1859–1869. https://doi.org/10.1098/rstb.2010.0398

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment.* https://www.coe.int/en/web/common-european-framework-reference-languages

Darragh, M., Booth, R. J., & Consedine, N. S. (2014). Investigating the ‘placebo personality’ outside the pain paradigm. *Journal of Psychosomatic Research*, *76*(5), 414–421. https://doi.org/10.1016/j.jpsychores.2014.02.011

Emadi Andani, M., Mahdavi, A., Bahrami, F., & Fiorio, M. (2015). Placebo-induced changes in behavioural parameters during gait: a pilot study. *2015 22nd Iranian Conference on Biomedical Engineering (ICBME)*, 337–340. https://doi.org/10.1109/ICBME.2015.7404166

Finniss, D. G. (2018). Placebo effects: Historical and modern evaluation. In *International Review of Neurobiology* (Vol. 139, pp. 1–27). Elsevier. https://doi.org/10.1016/bs.irn.2018.07.010

Fiorio, M., Andani, M. E., Marotta, A., Classen, J., & Tinazzi, M. (2014). Placebo-Induced Changes in Excitatory and Inhibitory Corticospinal Circuits during Motor Performance. *Journal of Neuroscience*, *34*(11), 3993–4005. https://doi.org/10.1523/JNEUROSCI.3931-13.2014

Ketsa (2021). *Bashful.* Retrieved 2 January 2022, from https://freemusicarchive.org/music/Ketsa/abundance/bashful

Geers, A. L., Kosbab, K., Helfer, S. G., Weiland, P. E., & Wellman, J. A. (2007). Further evidence for individual differences in placebo responding: An interactionist perspective. *Journal of Psychosomatic Research*, *62*(5), 563–570. https://doi.org/10.1016/j.jpsychores.2006.12.005

Geers, A. L., Wellman, J. A., Fowler, S. L., Helfer, S. G., & France, C. R. (2010). Dispositional optimism predicts placebo analgesia. *The Journal of Pain*, *11*(11), 1165–1171. https://doi.org/10.1016/j.jpain.2010.02.014

Glaesmer, H., Rief, W., Martin, A., Mewes, R., Brähler, E., Zenger, M., & Hinz, A. (2012). Psychometric properties and population-based norms of the Life Orientation Test Revised (LOT-R): Psychometric properties of the LOT-R. *British Journal of Health Psychology*, *17*(2), 432–445. https://doi.org/10.1111/j.2044-8287.2011.02046.x

Gracely, RichardH., Dubner, R., Deeter, WilliamR., & Wolskee, PatriciaJ. (1985). Clinicians' expectations influence placebo analgesia. *The Lancet*, *325*(8419), 43. https://doi.org/10.1016/S0140-6736(85)90984-5

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *11*(3), 501–518. https://doi.org/10.1037//0278-7393.11.3.501

Greifeneder, R., Bless, H., & Pham, M. T. (2011). When Do People Rely on Affective and Cognitive Feelings in Judgment? A Review. *Personality and Social Psychology Review*, *15*(2), 107–141. https://doi.org/10.1177/1088868310367640

Hall, K. T., Loscalzo, J., & Kaptchuk, T. J. (2015). Genetics and the placebo effect: the placebome. *Trends in Molecular Medicine*, *21*(5), 285–294. https://doi.org/10.1016/j.molmed.2015.02.009

Herrnstein, R. J. (1962). Placebo Effect in the Rat. *Science*, *138*(3541), 677–678. http://www.jstor.org/stable/1709743

Hinz, A., Sander, C., Glaesmer, H., Brähler, E., Zenger, M., Hilbert, A., & Kocalevent, R.-D. (2017). Optimism and pessimism in the general population: Psychometric properties of the Life Orientation Test (LOT-R). *International Journal of Clinical and Health Psychology*, *17*(2), 161–170. https://doi.org/10.1016/j.ijchp.2017.02.003

Hurst, P., Schipof-Godart, L., Szabo, A., Raglin, J., Hettinga, F., Roelands, B., Lane, A., Foad, A., Coleman, D., & Beedie, C. (2019). The Placebo and Nocebo effect on sports performance: A systematic review. *European Journal of Sport Science*, *20*(3), 279–292. https://doi.org/10.1080/17461391.2019.1655098

Jensen, K. B., Kaptchuk, T. J., Chen, X., Kirsch, I., Ingvar, M., Gollub, R. L., & Kong, J. (2014). A neural mechanism for nonconscious activation of conditioned placebo and nocebo responses. *Cerebral Cortex*, *25*(10), 3903–3910. https://doi.org/10.1093/cercor/bhu275

Jensen, K. B., Kaptchuk, T. J., Kirsch, I., Raicek, J., Lindstrom, K. M., Berna, C., Gollub, R. L., Ingvar, M., & Kong, J. (2012). Nonconscious activation of placebo and nocebo pain responses. *Proceedings of the National Academy of Sciences*, *109*(39), 15959–15964. https://doi.org/10.1073/pnas.1202056109

Jensen, K., Kirsch, I., Odmalm, S., Kaptchuk, T. J., & Ingvar, M. (2015). Classical conditioning of analgesic and hyperalgesic pain responses without conscious awareness. *Proceedings of the National Academy of Sciences*, *112*(25), 7863–7867. https://doi.org/10.1073/pnas.1504567112

Kaptchuk, T. J. (2018). Open-Label Placebo: Reflections on a Research Agenda. *Perspectives in Biology and Medicine*, *61*(3), 311–334. https://doi.org/10.1353/pbm.2018.0045

Kaptchuk, T. J., Kelley, J. M., Conboy, L. A., Davis, R. B., Kerr, C. E., Jacobson, E. E., Kirsch, I., Schyner, R. N., Nam, B. H., Nguyen, L. T., Park, M., Rivers, A. L., McManus, C., Kokkotou, E., Drossman,

D. A., Goldman, P., & Lembo, A. J. (2008). Components of placebo effect: randomised controlled trial in patients with irritable bowel syndrome. *BMJ*, *336*(7651), 999–1003. https://doi.org/10.1136/bmj.39524.439618.25

Keefe, R. S. E., Davis, V. G., Harvey, P. D., Atkins, A. S., Haig, G. M., Hagino, O., Marder, S., Hilt, D. C., & Umbricht, D. (2017). Placebo Response and Practice Effects in Schizophrenia Cognition Trials. *JAMA Psychiatry*, *74*(8), 807–814. https://doi.org/10.1001/jamapsychiatry.2017.1574

Keitel, A., Ferrea, S., Südmeyer, M., Schnitzler, A., & Wojtecki, L. (2013). Expectation Modulates the Effect of Deep Brain Stimulation on Motor and Cognitive Function in Tremor-Dominant Parkinson's Disease. *PLOS ONE*, *8*(12), e81878. https://doi.org/10.1371/journal.pone.0081878

Keitel, A., Wojtecki, L., Hirschmann, J., Hartmann, C. J., Ferrea, S., Südmeyer, M., & Schnitzler, A. (2013). Motor and cognitive placebo-/nocebo-responses in Parkinson's disease patients with deep brain stimulation. *Behavioural Brain Research*, *250*, 199–205. https://doi.org/10.1016/j.bbr.2013.04.051

Kern, A., Kramm, C., Witt, C. M., & Barth, J. (2020). The influence of personality traits on the placebo/nocebo response. *Journal of Psychosomatic Research*, *128*, 109866. https://doi.org/10.1016/j.jpsychores.2019.109866

Kirsch, I. (2018). Chapter Five - Response Expectancy and the Placebo Effect. In L. Colloca (Ed.), *International Review of Neurobiology* (Vol. 138, pp. 81–93). Academic Press. https://doi.org/10.1016/bs.irn.2018.01.003

Klinger, R., Stuhlreyer, J., Schwartz, M., Schmitz, J., & Colloca, L. (2018). Clinical use of placebo effects in patients with pain disorders. In *International Review of Neurobiology* (Vol. 139, pp. 107–128). Elsevier. https://doi.org/10.1016/bs.irn.2018.07.015

Kouider, S., & Dupoux, E. (2005). Subliminal speech priming. *Psychological Science*, *16*(8), 617–625. https://doi.org/10.1111/j.1467-9280.2005.01584.x

Lange, K., Kühn, S., & Filevich, E. (2015). "Just Another Tool for Online Studies" (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLOS ONE*, *10*(6), e0130834. https://doi.org/10.1371/journal.pone.0130834

Lowery, B. S., Eisenberger, N. I., Hardin, C. D., & Sinclair, S. (2007). Long-term effects of subliminal priming on academic performance. *Basic and Applied Social Psychology*, *29*(2), 151–157. https://doi.org/10.1080/01973530701331718

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Mikels, J. A., Maglio, S. J., Reed, A. E., & Kaplowitz, L. J. (2011). Should I go with my gut? Investigating the benefits of emotion-focused decision making. *Emotion*, *11*(4), 743–753. https://doi.org/10.1037/a0023986

Miller, F. G., & Kaptchuk, T. J. (2008). The power of context: reconceptualizing the placebo effect. *Journal of the Royal Society of Medicine*, *101*(5), 222–225. https://doi.org/10.1258/jrsm.2008.070466

Morton, D. L., Watson, A., El-Deredy, W., & Jones, A. K. P. (2009). Reproducibility of placebo analgesia: Effect of dispositional optimism: *Pain*, *146*(1), 194–198. https://doi.org/10.1016/j.pain.2009.07.026

Nuechterlein, K. H., Green, M. F., Kern, R. S., Baade, L. E., Barch, D. M., Cohen, J. D., Essock, S., Fenton, W. S., Frese, F. J., Gold, J. M., Goldberg, T., Heaton, R. K., Keefe, R. S. E., Kraemer, H., Mesholam-Gately, R., Seidman, L. J., Stover, E., Weinberger, D. R., Young, A. S., … Marder, S. R. (2008). The MATRICS Consensus Cognitive Battery, Part 1: Test Selection, Reliability, and Validity. *American Journal of Psychiatry*, *165*(2), 203–213. https://doi.org/10.1176/appi.ajp.2007.07010042

Petrie, K. J., & Rief, W. (2019). Psychobiological mechanisms of placebo and nocebo effects: Pathways to improve treatments and reduce side effects. *Annual Review of Psychology*, *70*(1), 599–625. https://doi.org/10.1146/annurev-psych-010418-102907

Pham, M. T., Lee, L., & Stephen, A. T. (2012). Feeling the Future: The Emotional Oracle Effect. *Journal of Consumer Research*, *39*(3), 461–477. https://doi.org/10.1086/663823

Popper, K. (1962). *Conjectures and refutations*. Routledge and Kegan Paul.

Rizzo, P. (2021). *Words and nonwords*. GitHub. https://github.com/Pietro-Rizzo/words_and_nonwords_1

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life orientation test. *Journal of Personality and Social Psychology*, *67*(6), 1063–1078. https://doi.org/10.1037/0022-3514.67.6.1063

Schmidt, L., Braun, E. K., Wager, T. D., & Shohamy, D. (2014). Mind matters: placebo enhances reward learning in Parkinson's disease. *Nature Neuroscience*, *17*(12), 1793–1797. https://doi.org/10.1038/nn.3842

Schwarz, K. A., & Büchel, C. (2015). Cognition and the Placebo Effect – Dissociating Subjective Perception and Actual Performance. *PLOS ONE*, *10*(7), e0130492. https://doi.org/10.1371/journal.pone.0130492

Spielberger, C., Gorsuch, R., Lushene, R., Vagg, P., & Jacobs, G. (1983). *Manual for the State-Trait Anxiety Inventory (Form Y1 – Y2): Vol. IV*.

Tulving, E. (1972). 10. Episodic and semantic memory. *Organization of memory/Eds E. Tulving, W. Donaldson, NY: Academic Press*, 381-403.

Turi, Z., Bjørkedal, E., Gunkel, L., Antal, A., Paulus, W., & Mittner, M. (2018). Evidence for cognitive placebo and nocebo effects in healthy individuals. *Scientific Reports*, *8*(1), 17443. https://doi.org/10.1038/s41598-018-35124-w

Turi, Z., Mittner, M., Paulus, W., & Antal, A. (2017). Placebo Intervention Enhances Reward Learning in Healthy Individuals. *Scientific Reports*, *7*(1), 41028. https://doi.org/10.1038/srep41028

VanArsdall, J. E., Nairne, J. S., Pandeirada, J. N. S., & Cogdill, M. (2015). Adaptive memory: Animacy effects persist in paired-associate learning. *Memory*, *23*(5), 657–663. https://doi.org/10.1080/09658211.2014.916304

Villa-Sánchez, B., Emadi Andani, M., Cesari, P., & Fiorio, M. (2021). The effect of motor and cognitive placebos on the serial reaction time task. *European Journal of Neuroscience, 53*(8), 2655–2668. https://doi.org/10.1111/ejn.15148

Wang, R.-S., Hall, K. T., Giulianini, F., Passow, D., Kaptchuk, T. J., & Loscalzo, J. (2017). Network analysis of the genomic basis of the placebo effect. *JCI Insight*, *2*(11), e93911. https://doi.org/10.1172/jci.insight.93911

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*(1), 23–28. https://doi.org/10.1080/17470218.2012.751117

Wei, H., Zhou, L., Zhang, H., Chen, J., Lu, X., & Hu, L. (2018, March 19). The Influence of Expectation on Nondeceptive Placebo and Nocebo Effects. *Pain Research and Management.* https://doi.org/10.1155/2018/8459429

Weimer, K., Colloca, L., & Enck, P. (2015). Placebo effects in psychiatry: mediators and moderators. *The Lancet. Psychiatry*, *2*(3), 246–257. https://doi.org/10.1016/S2215-0366(14)00092-3

Winkler, A., & Hermann, C. (2019). Placebo- and nocebo-effects in cognitive neuroenhancement: When expectation shapes perception. *Frontiers in Psychiatry*, *10*. https://doi.org/10.3389/fpsyt.2019.00498

Appendix A: Aftermath questionnaire

(The following instructions are displayed at the beginning of the questionnaire)

The following questions concern the Musical Pairs round you just completed. From your answers, I hope to get a better sense of your experience.

Please indicate to what extent you agree with the following statements, using the response format:

0 = strongly disagree, 1 = disagree, 2 = neutral, 3 = agree, 4 = strongly agree.

Please be accurate and honest in your answers.

(In random order, each item is then displayed individually, together with the digits from 0 to 4 displayed on a line, their respective meaning under each of them. Clicking any digit takes the participant to the next item.)

1. During Musical Pairs, I put forward my best effort.
2. During Musical Pairs, I felt absorbed.
3. During Musical Pairs, I was well-focused.
4. During Musical Pairs, I had some unrelated thoughts.
5. During Musical Pairs, I was interrupted.
6. During Musical Pairs, I felt tired.
7. During Musical Pairs, the environment around me was quiet.
8. During Musical Pairs, some noise around me made it harder for me to understand the words.
9. During Musical Pairs, I kept my eyes closed whenever either words or music were played.
10. During Musical Pairs, through Step 1, I tried to visualise each pair as an image in my mind.
11. During Musical Pairs, as the music played, I just listened, thinking of nothing in particular.

Appendix B: Musical Pairs questionnaire

(The following instructions are displayed at the beginning of the questionnaire)

The following questions concern all the Musical Pairs rounds you had during this experiment.
From your answers I hope to get a better sense of the participant's experience, check that everything worked as intended and inspire future improvements on the experimental protocol.
Please indicate to what extent you agree with the following statements, using the response format:
0 = strongly disagree, 1 = disagree, 2 = neutral, 3 = agree, 4 = strongly agree.
Please be accurate and honest in your answers.

(In random order, each item is then displayed individually, together with the digits from 0 to 4 displayed on a line, their respective meaning under each of them. Clicking any digit takes the participant to the next item.)

1. During Musical Pairs, the sound was crisp.
2. During Musical Pairs, the words were loud enough.
3. During Musical Pairs the words sounded somewhat distorted, or metallic.
4. During Musical Pairs, the words played were words I am familiar with.
5. During Musical Pairs, some of the words were new to me.
6. During Musical Pairs, the words were pronounced clearly.
7. During Musical Pairs, the words were pronounced with a rather heavy accent.
8. During Musical Pairs, the way words were pronounced made them harder to understand.