Cognition 214 (2021) 104753

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

The placebo effect: To explore or to exploit?[☆]

Kirsten Barnes, PhD[a,*], Benjamin Margolin Rottman, PhD[b,1], Ben Colagiuri, PhD[a,1]

[a] University of Sydney, Australia
[b] The University of Pittsburgh, USA

ARTICLE INFO

Keywords:
Explore exploit
Decision making
Placebo effects
Pain
Placebo analgesia

ABSTRACT

How people choose between options with differing outcomes (explore-exploit) is a central question to understanding human behaviour. However, the standard explore-exploit paradigm relies on gamified tasks with low-stake outcomes. Consequently, little is known about decision making for biologically-relevant stimuli. Here, we combined placebo and explore-exploit paradigms to examine detection and selection of the most effective treatment in a pain model. During conditioning, where 'optimal' and 'suboptimal' sham-treatments were paired with a reduction in electrical pain stimulation, participants learnt which treatment most successfully reduced pain. Modelling participant responses revealed three important findings. First, participants' choices reflected both directed and random exploration. Second, expectancy modulated pain – indicative of recursive placebo effects. Third, individual differences in terms of expectancy during conditioning predicted placebo effects during a subsequent test phase. These findings reveal directed and random exploration when the outcome is biologically-relevant. Moreover, this research shows how placebo and explore-exploit literatures can be unified.

1. Introduction

Understanding how humans learn about options available in their environment and make choices to achieve desirable outcomes has become a key focus of psychological research spanning cognitive psychology, associative learning, machine learning, health psychology, and behavioral neuroscience (Addicott, Pearson, Sweitzer, Barack, & Platt, 2017; Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006; Gershman, 2018; Gittins, 1979; Navarro, Newell, & Schulze, 2016; Raisch, Birkinshaw, Probst, & Tushman, 2009; Rottman, 2016; Smith & Tushman, 2005; Wilson, Geana, White, Ludvig, & Cohen, 2014). One of the fundamental questions in this area, referred to as the 'explore-exploit tradeoff', concerns how individuals decide between sticking with (i.e. exploiting) options that they have already tested versus exploring other options with possibly different outcomes (e.g. Mehlhorn et al., 2015; Wilson et al., 2014). Such decisions are inherent in virtually every aspect of our lives, sometimes with life-threating consequences, for example, deciding which medical treatment to take when facing a major illness – the established treatment or the supposedly promising one that has recently entered the market. In parallel, research on the placebo effect has sought to understand how individuals' beliefs about a treatment influence their actual experiences with the treatment (for review, see: Benedetti, 2008; Colagiuri, Schenk, Kessler, Dorsey, & Colloca, 2015; Finniss, Kaptchuk, Miller, & Benedetti, 2010; Geuter, Koban, & Wager, 2017; Price, Finniss, & Benedetti, 2008). One of the strengths of research on the placebo effect is that it uses genuine outcomes, such as pain and nausea, and therefore has high external validity in terms of the outcomes that are being experienced. Biologically-relevant stimuli of this type, such as pain, have been employed in decision making and choice paradigms that explore how aversive reinforcers impact the learning and decision making process (e.g. Roy et al., 2014; Seymour, Daw, Roiser, Dayan, & Dolan, 2012; Wang, Lee, O'Doherty, Seymour, & Yoshida, 2018; Wimmer & Büchel, 2020). For example, Wang et al. (2018) have modelled whether habit and rule-based decision making systems are engaged to avoid pain, demonstrating greater fluidity in the switching of control between systems in the case of aversive stimuli. Similarly, Roy et al. (2014) have investigated brain activity associated with prediction error signals generated during the avoidance of one of two options resulting in a greater probability of pain. However, none of these studies are explicitly concerned with the development of placebo analgesia. Conversely, a review of the placebo literature reveals that research in this area has typically employed passive designs that do not involve

[☆] Data files and analysis source code for all results are available at the following link: https://osf.io/dzem4/.
[*] Corresponding author at: School of Psychology, A18, University of Sydney, NSW 2006, Australia.
E-mail address: kirsten.barnes@sydney.edu.au (K. Barnes).
[1] Equal senior authorship.

https://doi.org/10.1016/j.cognition.2021.104753
Received 20 August 2020; Received in revised form 21 April 2021; Accepted 26 April 2021
Available online 20 May 2021

choice. This is with the exception of a group of studies, reviewed in the sections below, that explore a single choice of treatment, or choice over timing of treatment (e.g. Bartley, Faasse, Horne, & Petrie, 2016; Brown, Fowler, Rasinski, Rose, & Geers, 2013; Geers et al., 2013; Geers, Rose, Fowler, & Brown, 2014; Rose, Geers, Fowler, & Rasinski, 2013; Rose, Geers, Rasinski, & Fowler, 2012; Tang, Geers, Barnes, & Colagiuri, 2019). Hence, there currently exists a gap in our knowledge regarding how the decision making process impacts the placebo effect for multiple treatments over time. This includes whether placebo analgesia develops when an individual explores or exploits different treatments on offer and whether expectancies generated as a result of choice influence the placebo effect when a treatment is no longer effective.

In contrast, the primary goal of research on the explore-exploit tradeoff is to understand how people choose among options over time (e.g. Gershman, 2018). As outlined above, while research exists that has employed reinforcement learning paradigms in pain (e.g. Roy et al., 2014; Wang et al., 2018), none have focused on testing the algorithms for search (e.g. the process through which individuals explore or exploit). As such, existing research on the explore-exploit tradeoff lacks ecological validity because it typically has involved artificial scenarios involving gamified points or small monetary rewards. The types of gamified points and small monetary rewards employed in typical explore-exploit paradigms are likely much less motivating and biologically relevant than outcomes like pain. Further, while the subjective value of points and monetary rewards may change, e.g. gain vs. loss framing and contrast effects (e.g. Tversky & Kahneman, 1981), such processes are distinct from the placebo effect, whereby expectancies directly change perceptual experiences, such as pain (Wiech, 2016; Wiech, Ploner, & Tracey, 2008). As a consequence, explore-exploit paradigms have not explored whether and how outcomes can be modulated though expectation-driven placebo processes.

In the current study, we aimed to integrate these two traditionally disparate research areas, by developing an exploit-explore placebo pain paradigm to understand how individuals make decisions about options that have genuine physiological consequences, and to understand how the placebo effect plays out in a dynamic explore-exploit situation.

1.1. Overview and standard paradigms

1.1.1. The placebo effect

The placebo effect occurs when treatment cues themselves influence patient outcomes (i.e., it is the belief that one is receiving a particular treatment that leads to the outcome rather than the efficacy of the treatment itself). Expectancy is considered a central mechanism of the placebo effect and can arise from both instruction and learning processes (including explicit instruction, social observation, and Pavlovian conditioning; Colagiuri, Schenk, et al., 2015; Price et al., 2008; Stewart-Williams & Podd, 2004). A large body of evidence has shown that placebo-induced expectancy can influence subjective, psychophysiological, and neurobiological outcomes (e.g. Benedetti, Amanzio, Baldi, Casadio, & Maggi, 1999; Colagiuri & Quinn, 2018; Geuter, Eippert, Hindi Attar, & Büchel, 2013; Jepma & Wager, 2015; Montgomery & Kirsch, 1997).

One of the most frequently employed paradigms for studying the placebo effect, and the one that we build upon here, involves experimentally-induced pain (Au Yeung, Colagiuri, Lovibond, & Colloca, 2014; Bąbel et al., 2017; Colagiuri & Quinn, 2018; Colloca & Benedetti, 2009; Colloca, Petrovic, Wager, Ingvar, & Benedetti, 2010; Colloca, Sigaudo, & Benedetti, 2008; Voudouris, Peck, & Coleman, 1985). As shown in Row 1 of Fig. 1, this paradigm typically involves two phases: conditioning and test. Participants are given a bogus treatment, such as a sham transcutaneous electrical nerve stimulation (TENS) device, and told it will reduce their pain. In the conditioning phase, the intensity of an electrical shock delivered on treatment trials (e.g. 'TENS active') is surreptitiously reduced compared to no-treatment trials (e.g. 'TENS inactive'). Participants therefore come to expect analgesia (pain relief) with treatment (i.e. supposed TENS activity). In the test phase, all shocks are delivered at equivalent intensity, irrespective of whether or not the treatment is applied. If participants continue to report less pain on treatment, relative to no-treatment trials, then this is considered evidence of a placebo effect. While this paradigm has yielded significant insight into the behavioral and neuropsychological underpinnings of the placebo effect, it is important to note that behaviour during the conditioning phase has seldom been used to predict the placebo effect at test and decision-making processes have been largely ignored.

1.1.2. The explore-exploit tradeoff

The explore-exploit tradeoff refers to situations in which an individual needs to choose between two or more options and learn from experience about the value and distribution of outcomes associated with their choices. It is generally expected that individuals transition from 'exploring' the options (testing multiple options in order to identify the optimal choice) to 'exploiting' this choice to maximize positive outcomes (Mehlhorn et al., 2015). The typical paradigm for an explore-exploit study, as shown in Row 2 Fig. 1, involves three cyclic steps. First, the individual makes a choice between two or more options. Second, the outcome of that choice is experienced. Third, the outcome is integrated into the individual's knowledge about the expected outcomes produced by each option. The cycle then starts over with the individual making another decision.

There are many important questions in the explore-exploit literature such as how people handle the explore-exploit tradeoff in dynamic[2] situations (e.g. Biele, Erev, & Ert, 2009; Daw et al., 2006; Navarro et al., 2016; Rottman, 2016; Yi, Steyvers, & Lee, 2009), how closely human performance matches optimal performance vs. heuristics (e.g. Lee, Zhang, Munro, & Steyvers, 2011; Steyvers, Lee, & Wagenmakers, 2009; Yi et al., 2009), or how people come to form expectations of future outcomes from past experiences and whether they are biased by individual past experiences (e.g. Bornstein, Khaw, Shohamy, & Daw, 2017). In the following sections, we focus on two as yet untested questions spanning the explore-exploit and placebo literatures. The first concerns the role of placebo effects during the explore-exploit task, and how this impacts the strength of the placebo effect at test (i.e. where participants no longer have control over which treatment they receive). The second concerns how people come to choose between different treatment options on offer.

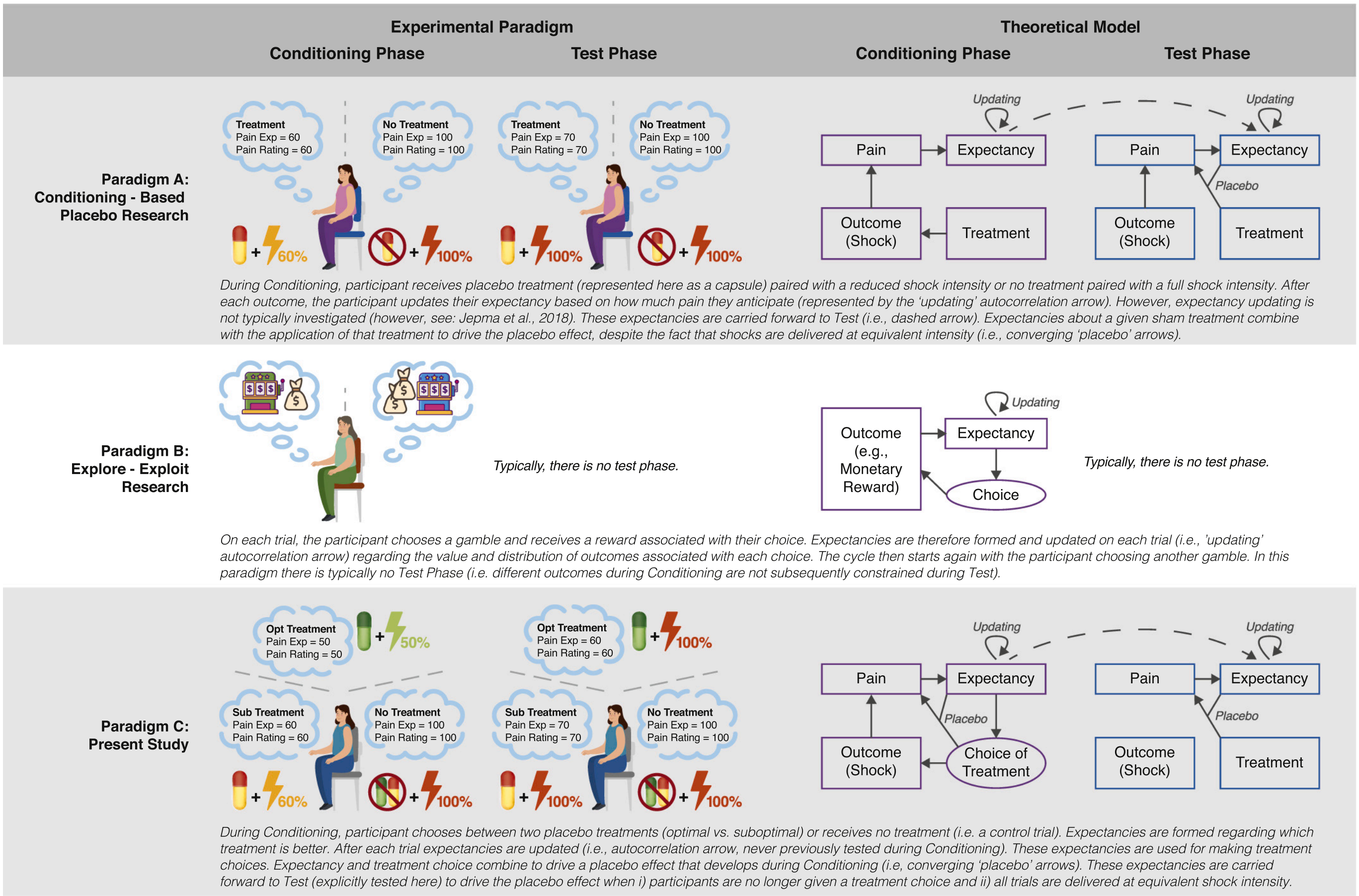
1.2. Question 1: placebo effects while navigating an explore-exploit challenge

One of the main goals of this research was to test for a cyclic process (expectations on pain, and pain on expectations) during the conditioning phase, which would provide evidence of a placebo effect during conditioning. Below we discuss the importance of this process within both placebo research and explore-exploit research. As the placebo literature has seldom focused on how behaviour occurring during conditioning phase can modulate the placebo effect at test, a secondary aim was to bridge this open question.

1.2.1. Placebo research

Existing placebo conditioning research has focused almost exclusively on the test phase. During this phase, the primary measure of the placebo effect is whether participants report different levels of pain for the treatment (e.g. 'TENS active') vs. no treatment (e.g. 'TENS inactive') conditions, despite the intensity of the shock (or other nociceptive

[2] Here we are focusing specifically on explore-exploit tasks that are 'static' in the sense that the outcomes associated with the choices are stationary over time. In dynamic explore-exploit tasks, the outcome associated with each choice can change over time, meaning that the learner must continually explore other options to see which option produces the best outcome *now*.


Experimental Paradigm
Conditioning Phase
Test Phase
Theoretical Model
Conditioning Phase
Test Phase
Paradigm A: Conditioning - Based Placebo Research
Treatment Pain Exp = 60 Pain Rating = 60
No Treatment Pain Exp = 100 Pain Rating = 100
Treatment Pain Exp = 70 Pain Rating = 70
No Treatment Pain Exp = 100 Pain Rating = 100
60%
100%
100%
100%
Updating
Pain
Expectancy
Outcome (Shock)
Treatment
Placebo
Paradigm B: Explore - Exploit Research
Typically, there is no test phase.
Outcome (e.g., Monetary Reward)
Expectancy
Choice
Typically, there is no test phase.
Paradigm C: Present Study
Opt Treatment Pain Exp = 50 Pain Rating = 50
50%
Opt Treatment Pain Exp = 60 Pain Rating = 60
100%
Sub Treatment Pain Exp = 60 Pain Rating = 60
No Treatment Pain Exp = 100 Pain Rating = 100
Sub Treatment Pain Exp = 70 Pain Rating = 70
No Treatment Pain Exp = 100 Pain Rating = 100
60%
100%
100%
100%
Choice of Treatment

During Conditioning, participant receives placebo treatment (represented here as a capsule) paired with a reduced shock intensity or no treatment paired with a full shock intensity. After each outcome, the participant updates their expectancy based on how much pain they anticipate (represented by the 'updating' autocorrelation arrow). However, expectancy updating is not typically investigated (however, see: Jepma et al., 2018). These expectancies are carried forward to Test (i.e., dashed arrow). Expectancies about a given sham treatment combine with the application of that treatment to drive the placebo effect, despite the fact that shocks are delivered at equivalent intensity (i.e., converging 'placebo' arrows).

On each trial, the participant chooses a gamble and receives a reward associated with their choice. Expectancies are therefore formed and updated on each trial (i.e., 'updating' autocorrelation arrow) regarding the value and distribution of outcomes associated with each choice. The cycle then starts again with the participant choosing another gamble. In this paradigm there is typically no Test Phase (i.e. different outcomes during Conditioning are not subsequently constrained during Test).

During Conditioning, participant chooses between two placebo treatments (optimal vs. suboptimal) or receives no treatment (i.e. a control trial). Expectancies are formed regarding which treatment is better. After each trial expectancies are updated (i.e., autocorrelation arrow, never previously tested during Conditioning). These expectancies are used for making treatment choices. Expectancy and treatment choice combine to drive a placebo effect that develops during Conditioning (i.e, converging 'placebo' arrows). These expectancies are carried forward to Test (explicitly tested here) to drive the placebo effect when i) participants are no longer given a treatment choice and ii) all trials are delivered at equivalent shock intensity.

Fig. 1. Comparison of typical conditioning-based placebo research (Paradigm A), typical explore-exploit research (Paradigm B), and the current study (Paradigm C), which is an amalgam of the two. The image depicts both the hypothesised theoretical underpinnings of each paradigm as well as the typical experimental design associated with each. We test the links outlined in our theoretical model (Paradigm C) in the results section and present a high-level summary of the findings in Fig. 6.

stimulus) delivered being equivalent. One recent study (Jepma, Koban, van Doorn, Jones, & Wager, 2018) analyzed the test phase in greater depth and demonstrated that, on a trial-by-trial basis, expectancy predicts subsequent pain, and the experienced pain predicts subsequent expectancy. This influence is depicted in the test phase of the conditioning-based placebo paradigm cell in Row 1 Fig. 1. This feedback loop is believed to explain the often-claimed persistence of the placebo effect even when – in the test phase – participants are receiving shocks of equal objective intensity (Anchisi & Zanon, 2015; Büchel, Geuter, Sprenger, & Eippert, 2014; Jepma et al., 2018; Wiech, 2016).

However, placebo effects do not only occur for inert treatments; they have also been demonstrated in the case of active treatments, such as enhancing the efficacy of morphine (Benedetti et al., 2003). As a consequence, the same recursive processes that drive the placebo effect during the test phase should also exist during the conditioning phase, when there really is a difference in the outcomes associated with the treatments. At present, however, such a feedback loop remains untested; thus, one main goal was to test for evidence of a placebo effect during conditioning.

Because this placebo effect process during conditioning has not been studied before, another goal was to test whether the magnitude of the placebo effect during conditioning predicted the magnitude of the standard way of measuring the placebo effect at test. If so, this would provide evidence of convergent validity, that this new way of measuring the placebo effect during learning is related to the standard measure of the placebo effect and is therefore likely to be tapping into the same underlying process.

1.2.2. *Explore-exploit research*

Another question in explore-exploit research, aside from choice, has to do with the role of expectations. Expectations are believed to play an critical role in how people learn to navigate the explore-exploit tradeoff. Reinforcement Learning (RL) models broadly assume that learners use their current expectations to make predictions about the future, and then update these expectations through an error-correcting learning algorithm, often called the Delta learning rule. Eq. (1) provides a simple version of the error correcting algorithm in the Rescorla-Wagner model (Rescorla, 1972), though the same fundamental idea applies to all RL models such as those cited later on in this paper. Q is the current associative strength, or in other words, the expectation. r is the outcome or experience on a given trial. α is the learning rate parameter, and ΔQ is the change in the associative strength due to the new experience.

$$\Delta Q = \alpha[r - Q] \tag{1}$$

RL models assume that the experienced outcomes are objective. However, in stark contrast, placebo-based research assumes that expectations bias the experience of outcomes. Imagine an extreme situation in which the placebo effect is so strong, that expectations bias the experiences so that the experiences are perceived as exactly in line with the expectations. In this case, the experience minus expectations term would be zero, producing no learning. Of course, we do not expect such an extreme placebo effect. Imagine instead that the experience (r) is perceived as the weighted average of the true experience (r) and the expectation (Q). We use p to denote the weighted average term – it stands for 'placebo'. When $p = 1$, the current experience is entirely driven by the prior expectation (Q), whereas when $p = 0$ there is no placebo effect and the experience is entirely determined by the objective stimulus (r). We can add this weighted version of the experience (λ) into Eq. (1), like in Eq. (2), which simplifies to Eq. (3).

$$\Delta Q = \alpha[((1-p)r + pQ) - Q] \tag{2}$$

$$\Delta Q = \alpha(1-p)[r - Q] \tag{3}$$

As can be seen in Eq. (3), when there is a complete placebo effect and the current experience is entirely driven by the prior expectation ($p = 1$) no learning occurs $\Delta Q = 0$. And when there is no placebo effect ($p = 0$), Eq. (3) is equivalent to Eq. (1). If there is an intermediary placebo effect (e.g., $p = .5$), then learning proceeds at a slower pace (.5α).

In summary, this analysis shows that the standard RL algorithm is in fact compatible with a placebo effect during learning whereby the placebo effect would slow down learning. This process of sequential learning with a placebo effect is depicted in Fig. 2.

However, research on the exploration-exploitation tradeoff (and RL more generally) typically involves gamified points or small monetary rewards. Such outcomes, presented as numbers, leave little room for bias by expectations, and in this setting the biasing role of prior expectations have rarely been studied (but see: Luhmann & Ahn, 2007, 2011; Marsh, 2009). However, many real-world outcomes such as physiological, emotional or behavioral outcomes are known to be modulated by expectancies (e.g. Tracey, 2010; Wiech, 2016). Therefore in the current research, we used pain as the outcome, which introduces the possibility of a placebo effect during an explore-exploit task, i.e. a placebo effect during learning.

1.3. Question 2: how people choose between options

1.3.1. Placebo effect research

At present, little is known about the development of the placebo effect in situations in which participants make choices about their treatment dynamically over time. As previously outlined, studies on the placebo effect that have involved choice have studied the role of making a single choice between two medications on the size of the placebo effect (e.g. Bartley et al., 2016; Brown et al., 2013; Geers et al., 2013; Geers et al., 2014; Rose et al., 2012; Rose et al., 2013), or have explored when to initiate a (placebo) treatment relative to no treatment (Tang et al., 2019). As a consequence, little is known about instances where participants have the opportunity to *decide* whether to initiate *different* treatments.

Understanding the role of choice in the placebo effect is critical from a practical perspective. First, for prescription medications, many patients choose not to take their medications as prescribed (Cutler & Everett, 2010; Osterberg & Blaschke, 2005). Second, for over the counter or ‘as needed’ medications that people self administer, little is known about how people choose between treatments. And third, the market for complementary and alternative medicine therapies, which individuals pay for despite minimal evidence of efficacy, is very large (Barnes, Bloom, Nahin, & Stussman, 2009). Understanding the role of choice in the placebo effect can help shed light on the recursive processes that are believed to unfold as people try out and evaluate therapies of varying efficacy (Rottman, Marcum, Thorpe, & Gellad, 2017).

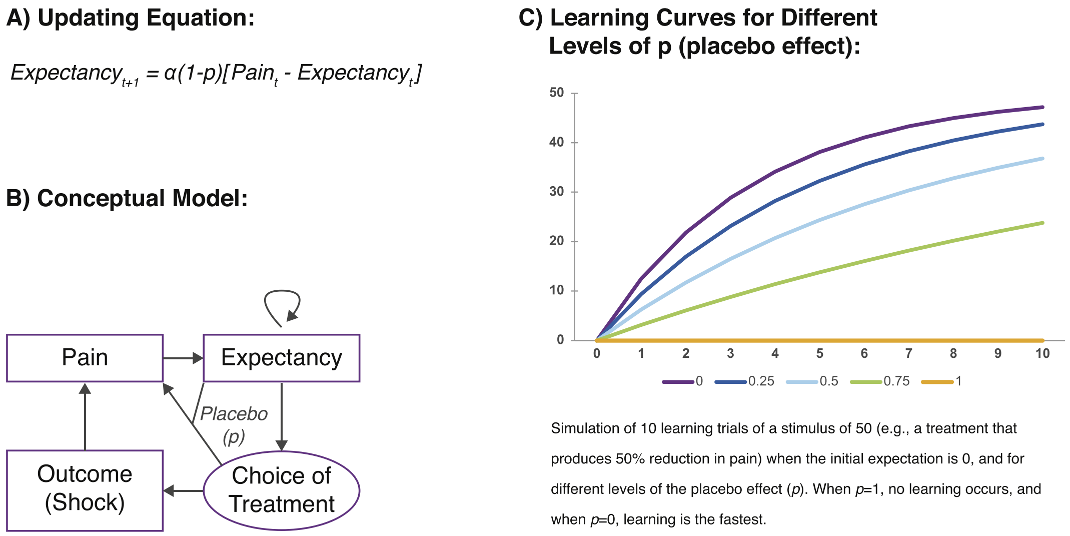
Understanding the role of choice in the placebo effect is also critical from a theoretical perspective. Research on illusory correlation shows that people can develop false beliefs about the efficacy of a treatment, which can be exaggerated when they have control (Blanco, Matute, & Vadillo, 2011; Bott & Meiser, 2020) and research on the placebo effect shows that expectations can shape participants’ experiences with a treatment. Together, this sets up the possbility of a feedback loop: expectations about treatments influence choices of treatments, choices of treatment influence an individual’s experiences, and these experiences then influence expectations.

In sum, placebo research provides little insight into choice and therefore lacks a connection to many real-world situations in which people make dynamic decisions intended to maximize the benefits and minimize the harms. In the current study, by modifying a placebo conditioning paradigm to allow participants to choose between treatments, we were able to use computational frameworks from the explore-exploit literature to understand how participants chose to test a given treatment at a given time.

1.3.2. Explore-exploit research

Gershman (2018, 2019) recently laid out an elegant way to analyze how people choose among options in explore-exploit tasks. This framework for analyzing choice is powerful because it cleanly distinguishes three distinct types of exploratory behaviour that map onto some of the most paradigmatic RL algorithms. Furthermore, each of these algorithms corresponds to easy to understand and compelling search behaviors.

The fundamental question addressed by these different search algorithms is whether and how the learner makes use of uncertainty about the outcomes of the available options when choosing among them (Gershman, 2018; Gershman & Tzovaras, 2018; Wilson et al., 2014). Here, we explain the theories in an intuitive and non-technical way. For technical details we refer the reader to the appendix, and also to



Simulation of 10 learning trials of a stimulus of 50 (e.g., a treatment that produces 50% reduction in pain) when the initial expectation is 0, and for different levels of the placebo effect (p). When p=1, no learning occurs, and when p=0, learning is the fastest.

Fig. 2. Updating equation, conceptual model, and learning curves, for a sequential model of learning with a placebo effect included.

Gershman (2018).

1.3.3. *Simple random exploration*

In random exploration algorithms, the choices do not depend on the relative uncertainty about the two options. According to the simplest type of random exploration, on each trial *t*, people tend to choose the option that they believe to be better; their choices in favor of one option over another are related to the difference in the expected rewards of the options, labelled V, which reflects an attempt to exploit the better option. Technically, the choice function uses the softmax function which, similar to logistic or probit regression, ensures that the probably of a given choice lies in the probability space 0 to 1 (e.g. Daw et al., 2006). Since choice behaviour is not deterministic, this is equivalent to saying that there a degree of random noise in this decision. This sort of exploration is called ‘random’; the learner tries to exploit the option that produces the better rewards, but also explores randomly to a certain extent.

1.3.4. *Thompson sampling*

Another, more sophisticated type of random exploration is called Thompson Sampling (Thompson, 1933). Thompson sampling draws a random sample from each of the posterior distributions for the available choices, and then picks the action with the highest sampled value (Russo, Roy, Kazerouni, Osband, & Wen, 2018). In the specific case in which there are only two choices and the posterior is Gaussian, as in the current study, the choice can be defined based on a ratio of V (the difference in the posterior value of the two options) over the *total uncertainty* (TU); V/TU (as derived in Gershman, 2018). For Gaussian distributions, TU is the square of the sum of the variances of the expectations of the two options. The variance of the expectation of each option decreases every time that option is chosen, thus, the most important feature of TU is that it decreases monotonically over the trials. Furthermore, it decreases faster at the beginning of learning.

For Thompson sampling, the ratio V/TU is the only predictor of choice. At the beginning of learning, when TU is high, V/TU is low, which predicts that the learner would not be able to distinguish between the two options well and choice would largely be random. As experience accumulates, TU decreases and V/TU increases, which predicts that the learner would have a stronger preference for the option with the better posterior mean (reflected in V). In sum, in simple random sampling even after many observations there is still randomness in every choice and the amount of randomness stays the same over time. In contrast, in Thompson sampling, the amount of randomness decreases as total uncertainty decreases – which happens with each additional sample that is experienced. This distinction is distilled in Situations 1 and 2 in Table 1.

1.3.5. *UCB*

A second set of theories involves ‘directed’ exploration; exploration is guided by the relative uncertainties about the two options. One type of directed exploration, embodied in the *Upper Confidence Bound (UCB;* Auer, Cesa-Bianchi, & Fischer, 2002*)* reinforcement learning algorithm, can be summarized as optimism in the face of uncertainty: the learner tends to choose options that they are more uncertain about (Relative Uncertainty; RU), in addition to preferring options that have better expected rewards (V). If the learner is more uncertain about Option A than Option B, but the average expectation of the two are the same, they will tend to pick A. This preference for ‘relative uncertainty’ is captured as the RU parameter in the choice model, which is simply the uncertainty of one choice minus the uncertainty of the other. Because the uncertainty of each option decreases each time it is chosen, RU is a function of the relative number of times that each option has been chosen.

Table 1
Summaries of the four sampling algorithms. The four rows provide and four situations that distinguish the four sampling algorithms. This table presents a situation in which there are two choices, A and B.

Situation	Simple (Random): V	Thompson (Random): V/TU	UCB (Directed): V + RU	Alternation (Directed): V - PC
1. After one choice of A and then one choice of B. Assuming that V somewhat favors A.	A > B due to V.	A > B due to V. TU is high.	A > B due to V. RU = 0 because both options chosen once.	A > B; alternate back to A.
2. Later in learning (e.g., after switching back and forth between A and B a number of times). Assuming that V somewhat favors A.	Same as above.	A> > B: stronger above because TU is now low.	Same as above.	Same as above.
3. After one choice of A, assuming V = 0, the second choice would tend to be:	A or B because V = 0.	A or B because V = 0 and therefore V/TU = 0.	B > A because RU is higher for B than A.	B > A because prior choice was A.
4. After choosing A, then B, assuming V = 0, the third choice would tend to be:	Same as above.	Same as above.	A or B because V = 0 and RU is the same for A and B since they were both chosen once.	A > B because the prior choice was B.

Note. V = difference in the posterior values of the two options. TU = total uncertainty. RU = difference in relative uncertainty for the two options. PC = prior choice.

For example, if a learner chooses A on Trial 1, the uncertainty will be much higher for B, leading to a strong preference for B on Trial 2. If the two options have been chosen the same number of times, there is no preference between the options. Near the end of learning, if Option A has been chosen one more time than Option B, Option B is slightly preferred for the next trial, but the strength of this preference is weaker than at the beginning of learning because uncertainty decreases faster at the beginning of learning. Because UCB is determined by both RU and V, and because the magnitude of RU decreases with experience, V comes to play a larger role over time. In sum, one way to think about UCB is that it predicts a higher rate of alternation between options at the beginning of learning.

Thompson sampling and UCB are similar in that Thompson sampling decreases exploration as total uncertainty decreases whereas for UCB sampling decreases exploration as relative uncertainty decreases. Therefore, both predict declining exploration and increasing exploitation of the better option over the course of learning. However, Gershman (2018, 2019) pointed out that these two strategies are behaviourally dissociable. In Thompson sampling high total uncertainty leads to more of a *random* choice, but in UCB high relative uncertainty leads to a *directed* choice to further test the option with higher uncertainty.

1.3.6. *Alternation vs. perseveration*

The prior three sections laid out reinforcement learning algorithms. Here we discuss a fourth tendency that can potentially influence participants’ choices; a habit to alternate. The degree of alternation (switching) vs. perseveration (making the same choice as the past choice) has in some literatures been used as a way to measure the degree of exploration vs. exploitation (Erev, 2010). In other research, the degree of alternation has been studied since exploring with high alternation rates can help the learner disconfound their choices from fluctuations in the background (Rottman, 2016). And, in yet other research the alternation/perseveration rate has not been of primary interest but has instead been a search habit that was statistically was controlled for when uncovering other search strategies (Bornstein et al., 2017). In fact, in most settings humans and non-human animals tend to perseverate, rather than alternate, and perseveration can be viewed as a way to minimize the complexity of action policies (Gershman, 2020).

Here we want to call attention to another feature of alternation: its place as a type of directed exploration and its relation to the UCB algorithm. Directed exploration is defined by a tendency to test options that have more uncertainty. In the framework above, the amount of uncertainty monotonically decreases with the number of times that an option has previously been tested. Strict alternation (in which sampling is only determined by the prior choice, not V and the prior choice) therefore ensures that the learner always chooses the option that has been tested the fewest number of times, minimizing RU. Though UCB also incorporates RU, when two options have been chosen the same number of times previously, the RU component of UCB is zero, so UCB would would choose randomly (so long as V = 0). In contrast, strict alternation predicts the learner to alternate between choices even when the two options have been chosen the same number of times. This distinction is distilled in Situations 3 and 4 in Table 1.

In sum, strict alternation effectively minimizes RU, so alternation could be viewed as a directed exploration strategy. Another interpretation of alternation is that it is a simple behavioral habit. Still, given that it happens to minimize RU for the Gaussian case, it can be viewed as a useful but simple heuristic. In our modelling, we used the subjects' prior choice (PC) as a predictor of their current choice. Alternation would be a negative influence, and perseveration would be a positive influence. Strict alternation would involve making choices simply based on the opposite of the prior choice (-PC). Alternatively and more realistically, search would also involve some degree of exploitation, so would also involve V in the form of V – PC.

1.3.7. *Summary of search algorithms*

Table 1 provides a summary of the four search algorithms and shows the relations between the search algorithms and the underlying components. The two directed algorithms are guided by relative uncertainty (RU) whereas the two random algorithms are not. The appendix provides more details about these search algorithms.

One of the main questions in this research is which search algorithm (s) best approximates human search. Prior research has found that a hybrid model including both components of Thompson sampling (random) as well as UCB (directed) best explain human choice patterns (Gershman, 2018, 2019; also see Gershman's, 2018 reanalysis of Wilson et al., 2014). Others have also found evidence for directed sampling based on relative uncertainty (Frank, Doll, Oas-Terpstra, & Moreno, 2009; Speekenbrink & Konstantinidis, 2015), though some have not (Daw et al., 2006). Here we test this same question, in a much more high stakes and physiologically relevant setting. In addition, we also test whether alternation is also a component of search, above and beyond the RU component of UCB. If so, this would suggest two types of directed search.

1.4. *The current study*

Row 3 in Fig. 1 represents the proposed model of the current study, which integrates the placebo effect paradigm and the explore-exploit paradigm. In our exploit-explore placebo paradigm, as with traditional placebo paradigms, both the conditioning and test phase contained treatment and no-treatment trials. Critically, however, on treatment trials during the conditioning phase, participants were offered the choice of two different (sham) TENS treatments which they were told would decrease their pain. One of these was 'optimal' in the sense that it resulted in a larger surreptitious reduction in shock intensity, making it appear more effective, than the 'suboptimal' treatment. It should be noted that the 'optimal' and 'suboptimal' labels were not employed in the study. Instead the TENS treatments were labelled as 'bipolar' and 'monopolar' (counterbalanced), with the participant having to discover which was the more effective. On no-treatment trials the TENS device was inactive and the shock was delivered at full intensity – considerably higher than either of the TENS treatment choices – making it seem as if both TENS treatments were better than no-treatment. During the test phase, participants were simply informed, without choice, about the trial type (i.e. 'monopolar', 'bipolar', 'no treatment') and all shocks were delivered at full intensity.

We investigated two main questions. First, Row 3 Fig. 1 Predicts a cyclic process (Pain ⟶ Expectancy ⟶ Pain). In particular the Expectancy ⟶ Pain link is evidence of a placebo effect. We test, for the first time, if this placebo effect is found during conditioning. This would also be the first evidence of the placebo effect occurring in an explore-exploit setting, requiring a reconceptualization of the error-driven learning processes underlying RL.

Second, based on the explore-exploit literature, we predicted a cyclic loop (Choice ⟶ Shock ⟶ Pain ⟶ Expectancy ⟶ Choice). We model each step along this process, but most importantly, we modelled choice using Gershman's framework explained above. Identifying the algorithms by which people make choices will generalize the results from prior explore-exploit tasks to a more authentic setting, and will mark the first result of this kind in a study on pain and the placebo effect.

In addition to these main questions, we also characterized participants' experiences of pain, expectancy, autonomic arousal, and choice, over the both the conditioning and test phases to understand the learning processes going on during the study; these analyses provide context for the main questions above.

2. Methods

2.1. *Participants*

Sixty-two undergraduate students were recruited from the University of Sydney (52% female, mean age = 19.4 years, ±2.2 *SD*). Of these, 57 participated for course-credit, while the remaining five were recruited from the general University population via a volunteer website and reimbursed AUD15. Participants were ≥18 years of age, fluent in English, were not currently experiencing pain or taking analgesic medication, had no history of heart problems, and had not participated in a similar study. The study was approved by the University of Sydney Human Research Ethics Committee.

2.2. *Cover story*

Participants were recruited under the pretense of a trial investigating the analgesic properties of transcutaneous electrical nerve stimulation (TENS) on pain. They were told that TENS involves passing a high-frequency, low-voltage current through the skin that stimulates the nerves to reduce pain (for example, see: Au Yeung et al., 2014). Participants were informed that two (bogus) types of TENS ('monopolar' and 'bipolar') were being investigated and that initial evidence suggests some individuals benefit from monopolar TENS, while others find bipolar TENS more effective. Critically, participants were informed that the aim of the experiment was to discover which type of TENS works best to reduce their pain. In reality, sham TENS stimulation was employed meaning no participant received any form of real TENS or any other active pain relief.

2.3. *Design and procedure*

Table 2 shows a summary of the study design, with the primary manipulation being within-participants comparison of the two sham treatments (optimal and suboptimal) relative to no-treatment and each other. In the conditioning phase, participants had the opportunity to

Table 2
Summary of experimental design (*Nb.* Percentages relate to the proportion of the participant's pain tolerance level at which shocks were delivered).

Conditioning *(3 Blocks of 10 trials)*	Test *(10 triads of 3 trials)*
2× No-Treatment (100%)	1× No-Treatment (100%)
8× Choice of Treatment:	1× 'Optimal' Treatment (100%)
-Optimal Treatment (50%)	1× 'Suboptimal' Treatment (100%)
-Suboptimal Treatment (60%)	

learn about the supposed efficacy (reduced pain) of the sham treatments relative to no-treatment. This involved 3 blocks of 10 trials. In each block, on 2 randomly determined 'non-treatment' trials participants received a shock at 100% of their pain tolerance threshold, providing a comparative baseline against sham treatment. For the remaining 8 trials per block (24 trials overall), participants choose between the two supposed treatments labelled 'monopolar' and 'bipolar'. These two 'treatment' labels were randomly assigned to either produce the greatest reduction in pain ('optimal treatment'; $M = 50\%$ of pain tolerance) or a smaller reduction in pain ('suboptimal treatment'; $M = 60\%$ of pain tolerance). This design makes it *appear* as if the sham TENS causes a reduction of pain relative to no-treatment, allowing participants to learn that both treatments are effective at relieving pain relative to no-treatment and that one treatment ('optimal') is more effective than the other ('suboptimal').

The test phase was initiated immediately following conditioning, with no instruction, signal, nor break indicating a change in phase. However, change in incidental cues (i.e. the omission of choice) would have alerted participants to a difference in procedure. The test phase consisted of 10 triads of 3 trials. During this phase, participants could not choose the sham treatments, and within each triad of 3 trials, there was one no-treatment trial, one optimal trial, and one suboptimal trial, randomly ordered. All shocks were delivered at 100% pain tolerance irrespective of treatment trial.

As part of the study, we also manipulated treatment variability by varying shock intensity in a between-groups design. Briefly, in the 'constant' condition, each treatment led to a constant level of shock reduction, while in the 'variable' condition the level of shock associated with a particular trial type varied around the mean value for that type. We hypothesised that the Variable group might better reflect real-world scenarios in which improvement in outcomes often fluctuates over time. However, as there were minimal differences between the two groups on any of the outcome measures, details regarding this manipulation are included as supplementary materials for the sake of brevity.

2.4. Procedure within each trial

For No TENS trials during the conditioning phase, participants were presented with the text "*Control trial. No TENS*" on the computer screen to denote trial type. As depicted in Fig. 3, on sham treatment trials, participants could choose between 'monopolar TENS' or 'bipolar TENS' by pressing the 2 or 3 key on the computer keyboard (counterbalanced). This decision was not time-limited. Subsequently, a 10-s countdown was presented. On sham treatment trials, TENS stimulation commenced after one second had elapsed and terminated at the end of the countdown. On all trials, during seconds 3 to 7 of the countdown, participants were prompted to rate their expectancy for pain via a response meter (model MLT1601/ST; ADInstruments) that used a continuous scale, with anchors marked at 0 (not), 50 (moderately), 100 (very). The countdown was followed by a fixation, 'X', on the screen for 0.5 s, which terminated with the shock. Participants subsequently rated their pain on a continuous computerized graphic rating scale from 0 (no pain) to 100 (very painful) using a mouse (as used by: Colagiuri & Quinn, 2018; Colagiuri, Quinn, & Colloca, 2015). Pain ratings were followed by a variable ITI of 10–15 s.

2.5. Apparatus

2.5.1. The 'TENS' device

The dummy TENS device consisted of a sham electrode attached to a stimulus isolator (Model FE180; ADInstruments, Bella Vista, NSW, Australia), and has been used frequently in previous research (e.g.Au Yeung et al., 2014 ; Colagiuri & Quinn, 2018 ; Tang et al., 2019). This was attached to the dorsum of the participant's non-dominant forearm and produced light tactile vibration and a beeping sound when supposedly active.

One TENS treatment was paired with a rapid activation of the electrode, the other was paired with a decelerated version of the same intensity (TENS stimulation: rapid vs. slowed). Slowed stimulation comprised17 x 2 Hz pulses of the TENS device. During rapid stimulation, a triplet of 5 Hz pulses was triggered every 1000 ms (pulse duration 400 ms, 600 ms interval). TENS stimulation of this type has repeatedly been shown to have no unconditioned effect on pain (Au Yeung et al., 2014; Colagiuri & Quinn, 2018; Colagiuri et al., 2015; Tang et al., 2019). Calibration was run for both types of TENS to ensure that participants could discriminate between them, with the experimenter calibrating the tactile stimulation for each participant between 2 and 3 mA as needed for discrimination.

2.5.2. Pain stimuli

Pain was induced by electrocutaneous stimulation (see: Au Yeung

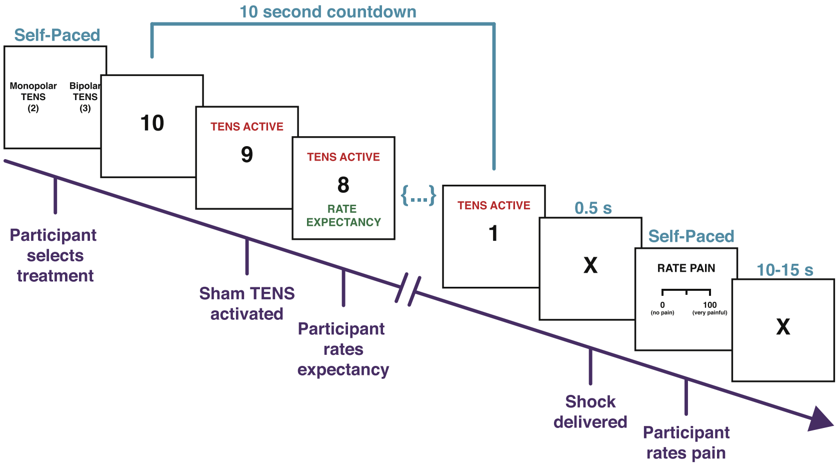


Fig. 3. Graphical representation of what was presented on the screen during the experiment. On sham treatment trials, the participant chose either monopolar or bipolar TENS. This initiated a 10 s countdown, with the sham TENS activated at second 3 of the countdown and expectancies rated (via response meter) from seconds 3–7. After a fixation (0.5 s) the shock was delivered and the participant rated their pain. On control trials participants were presented with a screen that read "*Control trial. No TENS*" which did not require choice. No choice was available during test. Instead, participants were presented with a screen informing them which sham treatment had been selected.

et al., 2014; Colagiuri & Quinn, 2018; Colagiuri, Quinn, & Colloca, 2015). Electric shocks were delivered via two Ag/AgCl electrodes connected to a stimulator (Model SHK1; Contact Precision Instruments, Cambridge, MA). Electrodes were attached to the dorsum of the non-dominant hand, approximately 1 cm apart. Shock stimuli consisted of 100-µs square pulses with a frequency of 100 Hz. The total duration of each shock was 0.5 s.

Intensity of the pain stimuli was calibrated for each participant in a stepwise fashion (as outlined in: Au Yeung et al., 2014; Colagiuri & Quinn, 2018; Colagiuri, Quinn, & Colloca, 2015). This was achieved by incrementing a low-level shock-intensity until the pain-stimulus was perceived as *painful, but tolerable*. Once this level was determined, participants were asked to rate their pain on a scale ranging from 0 (no pain) to 10 (very painful). If participants reported a value less than 6, they were asked if they would be comfortable sampling a higher intensity shock. The 100% shock intensity set for the present experiment therefore obtained a subjective pain-rating ≥ 6 across participants.

2.5.3. SCR

Skin conductance was recorded using two Ag/AgCl electrodes were attached to the distal phalange on the index and middle fingers of the non-dominant hand. The data were digitized at 1000 Hz using a PowerLab DAQ device coupled with a Galvanic Skin Response amplifier (ADInstruments, Bella Vista, NSW, Australia) and stored on a computer using Labchart software (Version 7, ADInstruments). Skin conductance data were downsampled to 50 Hz for analysis and then log transformed to correct for skew (see: Colagiuri & Quinn, 2018). Two target periods were defined: 1) the peak amplitude in anticipation of pain (6 s prior to shock-onset); 2) the peak amplitude of the phasic response to pain (6 s post shock-offset). Both were baseline-corrected to the averaged response in the last 2 s of the ITI to account for drift. Due to technical failure, three participants had incomplete skin conductance data. These participants are removed from all analyses concerning autonomic arousal.

3. Results

Data files and analysis source code for all the results are available at the following link: https://osf.io/dzem4/

3.1. *Univariate analysis of pain, expectancy, arousal, and choice during the conditioning phase*

During the conditioning phase, because participants could choose between the two sham treatments (except on no-treatment trials), there were variable numbers of observations of each treatment. For the pain, expectancy and skin conductance data, within-subjects ANOVAs with three levels (treatment: optimal, suboptimal, no-treatment) were run with contrast analysis employed to test for differences between the overall placebo effect (treatment vs. no-treatment) and the differential placebo effect (optimal vs. suboptimal treatment). This is consistent with typical analysis in the placebo literature (Au Yeung et al., 2014; Colagiuri & Quinn, 2018; Tang et al., 2019). Trial data were collapsed across conditioning as discrimination should be possible within the first few trials. The test was two-tailed, as were all others outlined below.

Fig. 4 shows mean average pain ratings, expectancy ratings and skin conductance in the conditioning phase. We first assessed whether participants could discriminate between the optimal and suboptimal sham treatment relative to no-treatment during conditioning, which were paired with objectively different shock levels. As expected, participants reported less pain after receiving any sham treatment relative to no-treatment $F(1,61) = 377.07$, $p < .001$, $\eta_p^2 = 0.86$, as well as after

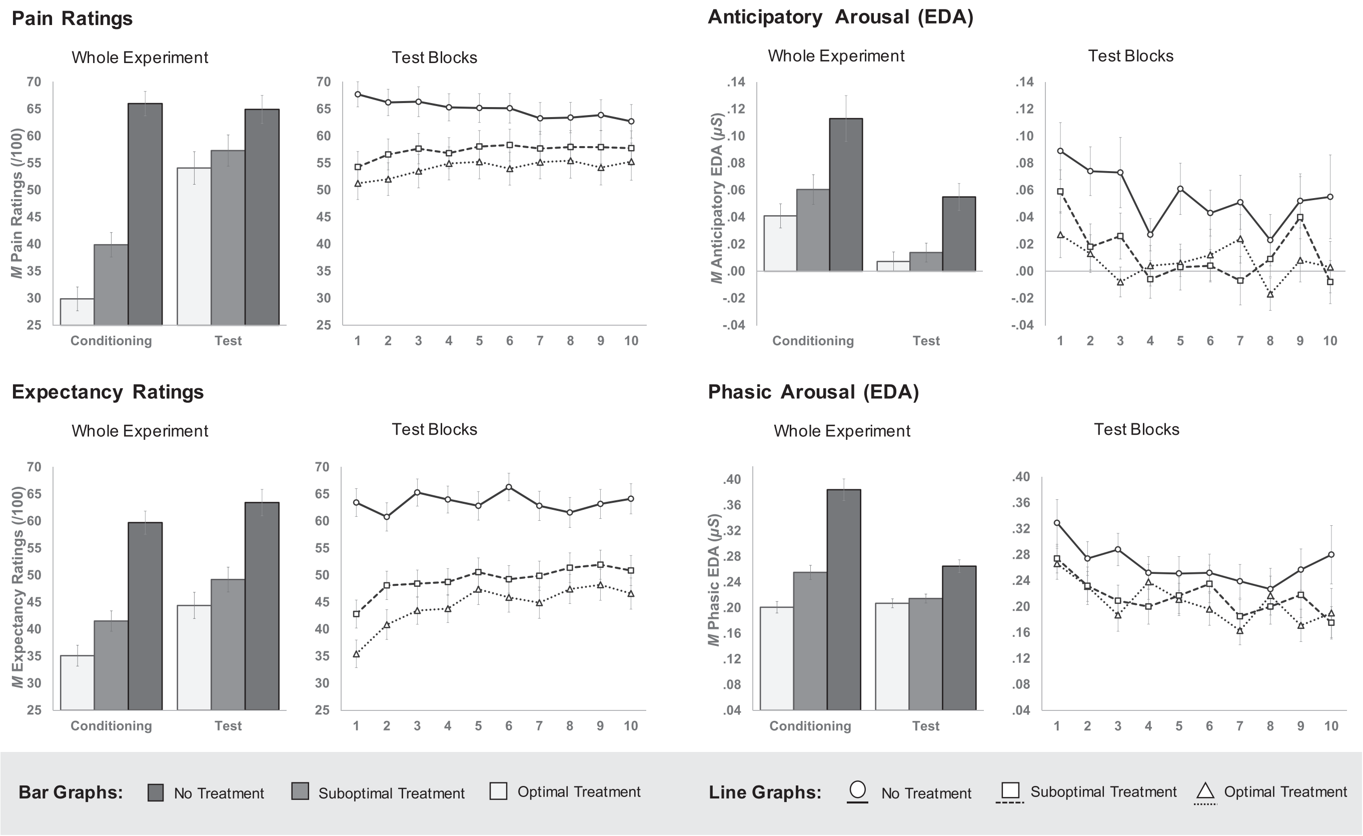


Fig. 4. Summary of univariate analyses. All 30 trials of conditioning and test are averaged in the whole experiment graphs. In the test phase, the 30 trials are aggregated into the 10 blocks associated with each sham treatment. Participants could only choose treatment options in the conditioning phase. All error bars represent the standard error of the mean (SEM).

choosing the optimal (largest shock reduction) relative to the suboptimal sham treatment (moderate shock reduction) $F(1,61) = 132.46$, $p < .001$, $\eta_p^2 = 0.69$. This confirmed that the shocks were discriminable and that the optimal treatment could be exploited to minimize the pain experienced.

Next, we were interested in whether participants expected a difference in pain to occur between the sham treatment choices, as this should be a precursor any placebo effect. As predicted, participants expected less pain on sham treatment relative to no-treatment trials $F(1,61) = 113.51$, $p < .001$, $\eta_p^2 = 0.65$, as well as after choosing the optimal relative to the suboptimal treatment $F(1,61) = 38.10$, $p < .001$, $\eta_p^2 = 0.38$. Anticipatory arousal, averaged across the conditioning phase, was found to be reduced in the six seconds prior to shock for both sham treatments relative to no-treatment $F(1,58) = 15.80$, $p < .001$, $\eta_p^2 = 0.21$, and was lower when the optimal treatment was chosen relative to suboptimal $F(1,58) = 4.64$, $p = .035$, $\eta_p^2 = 0.074$, indicating lower anticipatory autonomic arousal for sham treatments associated with less pain. Similarly, phasic arousal, averaged across the conditioning phase, was reduced for both treatments relative to no-treatment $F(1,58) = 89.18$, $p < .001$, $\eta_p^2 = 0.61$, and for the optimal treatment relative to the suboptimal $F(1,58) = 42.70$, $p < .001$, $\eta_p^2 = 0.42$.

In analyzing the choices between the two sham treatment types in order to assess whether participants had started to exploit to some extent during conditioning, the data were assessed in two ways; for both analyses, the 6 no-treatment trials were ignored. First, we assessed whether participants were more likely than chance to choose the optimal option within the last third of trials using a logistic regression with a by-subject random intercept. The last third of trials were selected for theoretical and practical reasons. Participants were required to learn which treatment was most effective and therefore any chance to exploit the optimal treatment should have emerged towards the end of the conditioning phase. As conditioning comprised three blocks of ten trials, splitting the conditioning phase into thirds evenly divided the available trials. By the third block of trials, participants were above chance in choosing the optimal treatment ($M = 0.58$, 95% $CI = [0.54 \text{ to } 0.63]$, $p < .001$), implying some amount of exploitation (Fig. 5). However, the fact that they were only slightly above chance, i.e. 58%, means that the amount of exploitation was not very high. Second, we assessed whether the rate of switching between the two sham treatments decreased over the 24 trials; as a participant becomes increasingly confident of which option is more effective they should switch less. We used a random effects logistic regression, with by-participant random intercepts and a random slope of trial number to account for repeated measures. The correlation between the random intercept and slope was dropped for convergence. We found that the switch rate decreased significantly across trials ($b = -0.048$, $SE = 0.011$, $p < .001$). However, despite the decrease, during the third block participants were still switching at a relatively high rate (60%), indicating that they were still engaging in some exploratory behaviour.

In summary, the univariate analysis of the conditioning phase revealed that participants learnt to distinguish the two treatment options from the no-treatment option and the optional from suboptimal treatment options. To this end, the participants started to exploit the optimal treatment; however, these changes over time were small. This could be a sign that discriminating between the efficacy of the two sham treatments was challenging. In addition, participants did not know that there were only 24 trials that they could choose, which may have suppressed exploiting if they thought that they would continue to explore for longer.

3.2. Univariate analysis of pain, expectancy, and arousal during the test phase

Analyses of pain, expectancy, and skin conductance were conducted to test for the presence of a placebo effect. The data were analyzed at the level of each triad (a triad consisted of one each of no-treatment, suboptimal, and optimal trials), with contrast analysis to test for an overall effect and a differential effect. Interactions with linear trends over time was used to test for differences in the rate of extinction of the placebo effect (i.e. analysis was conducted at the level of each trial over the test phase). Fig. 4 also shows mean pain, expectancy, and autonomic arousal during the test phase, both averaged and across individual trials. Treatment type was pre-determined in the test-phase to allow us to examine the placebo effect so there is no choice data in this phase.

Averaged across the test phase, participants reported lower pain during sham treatment trials than no-treatment trials, $F(1,61) = 58.83$, $p < .001$, $\eta_p^2 = 0.49$. As shock intensity was the same for sham treatment and no-treatment trials, this indicates there was a significant ‘overall’ (treatment vs. no-treatment) placebo effect. Further, participants reported lower pain during ‘optimal’ than ‘suboptimal’ treatment, $F(1,61) = 20.51$, $p < .001$, $\eta_p^2 = 0.25$, which we call a ‘differential’ placebo effect. Linear trend analyses across triads (i.e. over time) revealed that the linear trend interacted with sham treatment type (both treatments vs. no-treatment) suggesting that the overall placebo effect diminished over time, $F(1,61) = 25.38$, $p < .001$, $\eta_p^2 = 0.29$. The slopes for the optimal and suboptimal treatments were not significantly different, however, suggesting that the placebo effect for both sham treatments extinguished at a similar rate, $F(1,61) = 0.59$, $p = .445$, $\eta_p^2 = 0.01$.

As with the pain ratings, during the test phase, participants reported lower expectancy for pain on sham treatment relative to no-treatment trials, $F(1,61) = 89.89$, $p < .001$, $\eta_p^2 = 0.60$ and lower expectancy for pain on optimal versus suboptimal treatment trials, $F(1,61) = 19.21$, $p < .001$, $\eta_p^2 = 0.24$. Linear trend analysis also revealed an interaction

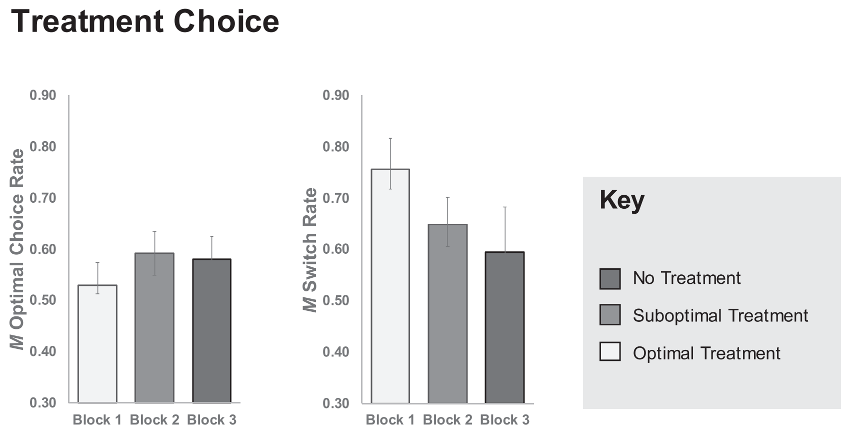


Fig. 5. Summary of sham treatment choice (optimal choice and switch rate). Given that analyses were logistic regressions, error bars are represented upper and lower 95% CIs rather than as the SEM (as depicted in Fig. 4).

with treatment type (treatment vs. no treatment) with expectancy for pain increasing over time on sham treatment trials, $F(1,61) = 10.22, p = .002, \eta_p^2 = 0.14$, indicating extinction of expectancy for treatment vs. no treatment. As with pain itself, the slopes for expectancy in optimal and suboptimal treatments were not significantly different, $F(1,61) = 1.92$, $p = .171, \eta_p^2 = 0.03$, suggesting a similar rate of extinction of expectancy for the two sham treatments.

Anticipatory arousal in the test phase was lower during sham treatment than no treatment trials, $F(1,58) = 45.61, p < .001, \eta_p^2 = 0.44$. This suggests that the expectation of pain reduction reduces autonomic arousal in anticipation of placebo treatment. No statistically significant difference in anticipatory arousal was observed between the two sham treatment types, $F(1,58) = 0.97$, $p = .329$, $\eta_p^2 = 0.02$. No interaction between sham treatment type and the linear trend over trials was observed for either the overall or differential placebo effect (all $ps >$ 0.380) suggesting the anticipatory arousal may have been more resistant to extinction than pain and expectancy.

Similar results were observed for the phasic response to the shock. Phasic arousal was lower during sham treatment trials compared to no treatment trials $F(1,58) = 37.18$, $p < .001$, $\eta_p^2 = 0.39$, demonstrating modulation of the autonomic response. No statistically significant difference in phasic arousal was observed between the two sham treatments $F(1,58) = 0.66, p = .422, \eta_p^2 = 0.01$. No interaction between sham treatment type and the linear trend over trials was observed for either the overall or differential placebo effect (all $ps >$ 0.640) suggesting limited extinction of this response.

In summary, consistent with previous research, univariate analysis of the test phase confirmed a placebo effect for pain for sham treatment vs. no treatment accompanied by greater expectancies for reduced pain as well as lower anticipatory and phasic autonomic arousal, for which previous evidence has been more mixed (Colagiuri & Quinn, 2018; Jepma & Wager, 2015). Importantly, the analysis also demonstrated that participants could learn about subtle differences between two effective treatments whereby the optimal treatment elicited a larger placebo effect for pain than did the suboptimal treatment. This was accompanied by a difference in expectancy, though not anticipatory or phasic arousal. When sham treatment was compared to no treatment, both pain and expectancy appeared to extinguish over the test phase while arousal did not, suggesting differences in the trajectory of self-report and physiological indices in response to and in anticipation of the placebo effect over time.

3.3. *Integrating pain, expectancy, and choice to test the theoretical model of exploring vs. exploiting*

Fig. 6 shows a summary of the results in the form of an integrated pain, expectancy, choice model supported by our data.

3.3.1. *Modelling pain ratings*

3.3.1.1. Pain ratings within the conditioning and test phases. The first question in terms of modelling pain ratings was whether a participant's expectancy on a given trial predicted their pain rating on that trial, as this would be evidence for a placebo effect both within the conditioning phase and test phase. In this regression, for the conditioning phase, we controlled for the intensity of the shock. We used multiple regression with by-participant random intercepts and random slopes for both predictors, and standardized all variables. For the test phase, all the shocks were the same so there is no need to statistically control for shock. A regression with a by-participant random intercept and random slope was used with standardized variables, dropping the correlation term between the intercepts and slopes for convergence. We used r^2_{NSJ} from the r2glmm package in R as the effect size measure, which is appropriate for random effects models (Johnson, 2014).

Table 3

Regressions predicting: 3A) pain during conditioning and test; 3B) the placebo effect during triad 1 of the test phase for the differential and overall placebo effect; 3C) expectancy during conditioning and test. 3A and 3C are random effects models and therefore have a different effect size ($r^2_{p\text{NSJ}}$) compared to 3B, which is a standard regression. All coefficients are standardized.

3A) Regressions predicting pain	Conditioning Phase		Test Phase	
Predictor:	*B*	$r^2_{p\text{NSJ}}$	*B*	$r^2_{p\text{NSJ}}$
Expectancy	0.29***	0.127	0.35***	0.163
Shock	0.44***	0.255	–	–

3B) Regressions Predicting Placebo Effect - Triad 1, Test Phase	Differential Placebo		Overall Placebo	
Predictor:	*B*	r^2_p	*B*	r^2_p
Expectancy Difference	0.45***	0.186	0.31**	0.118
Individual Difference Measure of Placebo Effect	0.12	0.014	0.43***	0.206
Exploitation – Percent Optimal Choices	−0.06	0.004	–	–
Exploitation – Switch Rate	0.03	0.001	–	–

3C) Regressions Predicting Expectancy	Conditioning Phase		Test Phase	
Predictor:	*B*	$r^2_{p\text{NSJ}}$	*B*	$r^2_{p\text{NSJ}}$
Optimal Choice (vs. No-TENS)	−0.29***	0.017	−0.28***	0.027
Suboptimal Choice (vs. No-TENS)	−0.23***	0.011	−0.22***	0.017
Prior Expectancy	0.33***	0.114	0.29***	0.084
Prior Pain	0.36***	0.111	0.35***	0.128

*$p < .05$.
**$p < .01$
***$p < .001$.

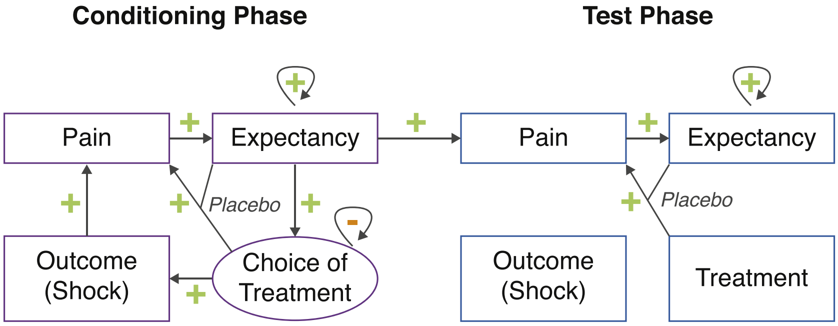


Fig. 6. High-level summary of results, where green '+' symbols represent theoretical links confirmed by our analyses, while orange '-' symbols represent statistically non-significant links. In the conditioning phase, the choice of treatment determines the shock, which influences the level of pain, which is aggregated into an expected level of pain for that sham treatment. The expectancies for the two treatments influence the choice of subsequent treatments, and the expectancy for a chosen treatment also recursively influences pain as a placebo effect. In the test phase, now all treatments have the same shock, and because the level of shock has zero variance, it is not connected statistically to any other variable. Also, participants no longer choose the treatment on a given trial. The pain is still aggregated into an expected level of pain for each treatment, and the expectancy for a given treatment recursively influences subsequent pain as a placebo effect. The expectancies at the end of the conditioning phase also influence the pain at the beginning of the test phase and expectancy at the beginning of the test phase. Please note that the paths represented in this figure are drawn from multiple analyses. As the model was not tested holistically, coefficients are not included. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3A presents the results of this analysis. As expected, the shock intensity on a given trial significantly predicted pain on that trial. However, most importantly, above and beyond shock, expectancy also significantly predicted participants' pain ratings, with a large effect size, which is evidence of a placebo effect occurring on a trial-by-trial basis. This finding held for both the conditioning and test phases. This provides evidence for a link between expectancy and pain in each phase of our experiment as proposed in our theoretical model in Fig. 1C.

3.3.1.2. Pain ratings across the conditioning and test phases. The second question regarding pain ratings was whether individual differences in the conditioning phase predict the placebo effect (i.e. the reduced perception of pain on treatment trials) during the first triad at test. Consistent with our previous research (e.g. Au Yeung et al., 2014; Colagiuri, Quinn, & Colloca, 2015), the first triad was chosen as it represents the strength of the placebo effect prior to any extinction having taken place (which was significant in this study). In this analysis we use conditioning phase data to predict the placebo effect (i.e. the reduced perception of pain on sham treatment trials) during test, both for the overall (treatment vs. no treatment) and for the differential placebo effect (optimal vs. suboptimal) via separate regressions. These regression results are presented in Table 3B.

The first predictor was the difference in expectancy for pain relief: the difference in expectancy for sham treatment vs. no-treatment for the overall analysis, and the difference in expectancy for optimal vs. suboptimal for the differential analysis) in the last third of conditioning trials. For the three participants who did not test the suboptimal choice in the last third, the most recent third in which the sham treatment was chosen was used instead. Theoretically, if participants come to expect a bigger difference during conditioning, that should predict a larger placebo effect at test. For both analyses, this effect was significant (Table 3B).

The second predictor was an individual difference measure of the size of the placebo effect *during conditioning*. The goal was to test whether the way of measuring the placebo effect in the prior section predicts this more traditional measure of the placebo effect. We extracted participant-level random effects of expectancy on pain during conditioning for the overall placebo effect (see Table 3A), and used the participant-level slopes to predict the placebo effect on the first test block. This individual difference measure was significant for the overall placebo effect, though not for the differential placebo effect (Table 3B). This provides some evidence that our analysis measuring the placebo effect during conditioning is predictive of the standard placebo effect measured at test (though only for treatment vs. no-treatment). In contrast, the amount of exploitation (optimal choice and switch rate) during the last third of conditioning does not predict the magnitude of the placebo effect (overall or differential) at the beginning of the test phase.

Finally, we tested whether participants who 'exploit' more might exhibit a larger differential placebo effect – as engaging in exploitation behaviour could be a proxy for believing in a difference in efficacy, similar to expectancy difference. This measure was captured with the percent of optimal choices and the switch rate in the last third of conditioning trials. However, according to the theoretical model, choice does *not* directly affect pain; expectancy does. Therefore, a non-significant effect when controlling for expectancy is consistent with our model, though we did not run Bayesian statistics to test for evidence for a null effect. Indeed, there was not a significant effect of either of these two measures of exploitation during conditioning on the placebo effect at test (Table 3B).

3.3.2. Modelling expectancy ratings

We next tested how participants update their expectancy ratings for a given treatment, based on their expectancy prior to the last time they experienced that sham treatment and their pain rating after experiencing it (Table 3C). This was done to provide evidence for a link between Pain and Expectancy within both phases of the theoretical model and provide evidence for a sequential learning process, represented by the circular Expectancy loops within each phase.

In this regression, all variables were standardized. By-subject random intercepts and random slopes for prior expectancy and prior pain were used. Expectancy ratings for all three choices (no-TENS, optimal, and suboptimal) were included in this regression, and there are fixed effects for the optimal and suboptimal choices (no-TENS was the reference group). Due to difficulties with convergence, by-subject random slopes were not included for these choices, and the correlations between the random slopes was also dropped for convergence.

As presented in Table 3C, participants' expectancy ratings for the optimal and the suboptimal options were lower than for no-treatment. More importantly, their expectancy ratings were positively related to their prior expectancy as well as their prior pain rating from their most recent experience of that option. These results mean that when providing an expectancy judgment for a particular choice (e.g., suboptimal), participants' expectancy rating is related to their most recent expectancy judgment for the same choice and their most recent experience of pain associated with that choice. This analysis is generally consistent with a wide class of standard reinforcement-learning error updating models in which expectancies are updated as a combination of the prior expectancy and the prior experience.

Finally, to confirm our theoretical model that predicts expectancies developed during conditioning will be carried forward to influence initial expectancies during test (dashed arrow, Fig. 1 Model C), we also directly correlated the difference in expectancy between the optimal and suboptimal treatment during the last third of conditioning with the difference in expectancy between treatments during the first triad to test. Results confirmed that expectancies developed during conditioning are associated with expectancies at the beginning of test, $r(60) = 0.46$, $p = .001$.

3.3.3. Modelling choice

Building on Gershman's (Gershman, 2018, 2019) approach, we tested 12 models of choice (Table 4). Models 7–12 parallel Models 1–6, except they also include a predictor of the prior choice (PC); if the influence of the prior choice on the subsequent choice is negative, it is a sign of alternation. Models 1 and 7 are the null model, which just has a by-subject random intercept. Within Models 2–6 and 8–12, we tested the three main models: the 'pure random' exploration model which only has V, the UCB model which has V and RU, and Thompson sampling, which has V/TU. In addition, we tested two other models. Mixture 2 has all three predictors from the Random, UCB, and Thompson models. Mixture 1 has RU and V/TU but not V; we found that V and V/TU were highly correlated and felt that Mixture 1 was sensible as an alternative because

Table 4
Model fit scores of 12 models of choice behaviour.

Number	Model Name	Predictors	AIC	BIC
1	Null	–	1955.3	1965.8
2	Pure Random	V	1947.8	1963.6
3	UCB	V + RU	1799.5	1825.8
4	Thompson	V/TU	1949.6	1965.4
5	Mixture 1	RU + V/TU	1798.5	1824.9
6	Mixture 2	V + RU + V/TU	1800.5	1832.1
7	Alt.	PC	1721.3	1742.3
8	Alt. + Pure Random	PC + V	1713.0	1739.3
9	**Alt. + UCB**	**PC + V + RU**	**1688.3**	**1725.2**
10	Alt. + Thompson	PC + V/TU	1712.8	1739.1
11	***Alt. + Mixture 1***	***PC + RU + V/TU***	***1687.5***	***1724.3***
12	**Alt. + Mixture 2**	**PC + V + RU + V/TU**	**1689.4**	**1731.5**

Note. V = difference in expected value. TU = total uncertainty. V/TU = V divided by TU. RU = relative uncertainty. PC = prior choice. Models in bold (9, 11, and 12) are the best fitting models and Model 11 (bold and italics) is the absolute best.

it includes both V (through V/TU) as well as the two types of directed sampling.

To conduct this analysis we ran a logistic regression with a probit link function with three predictors. The dependent variable was participants' choice: 1 was the optimal choice and 0 was the suboptimal choice. In line with his approach, we used the Kalman filter equations to calculate V, RU, and V/TU. For details see the appendix, and also see Gershman (2018, 2019).

Because of the repeated measures (24 choices per participant, minus the first for which they have no beliefs), we used a random effects model. We attempted to fit the maximal models with a random intercept and random slope for each of the predictors; however, we found that sometimes the models would not converge if there were random slopes for V and for V/TU. We found that by dropping the random slopes for V and V/TU and keeping the by-subject random intercepts and random slopes for PC and RU, all of the models would converge. This was done for all of the models, rather than on a model-by-model basis, to make the AIC and BIC scores more comparable by using the same random effects structure.

Technically, Model 11 (Alternation + Mixture 1) had the lowest AIC and BIC (Table 4). This model suggests that the participants explored in both directed (PC, RU) and random (V/TU) ways. The influence of prior choice (PC) was negative, which captures a high alternation rate between treatment options, $b = -0.92$, SE $= 0.12$, $p < .001$, $r^2_{p\mathrm{NSJ}} = 0.065$. The effect of relative uncertainty (RU) between the two choices was positive, $b = 0.25$, SE $= 0.07$, $p < .001$, $r^2_{p\mathrm{NSJ}} = 0.043$, and so was the effect of the difference in expected value of the two choices normalized using the total uncertainty (V/TU), $b = 0.26$, SE $= 0.09$, $p = .002$, $r^2_{p\mathrm{NSJ}} = 0.011$. However, some caution is required here, because when using Burnham and Anderson's (2004) criteria for interpreting AIC, there were limited differences if fit between Models 9, 11, and 12.

Still, stepping back and looking at patterns in all the models still produces a clear picture. First, it is very clear that all the models with alternation (Models 7–12), a type of directed sampling, fit much better than those without (1–6). Second, it is also very clear that the models with RU, another type of directed sampling, fit considerably better than similar models without RU; compare models 2 vs. 3, 4 vs. 5, 8 vs. 8, and 10 vs. 11. In fact, a random effect regression model with alternation and RU, and random slopes for both, reveals a correlation of $r = 0.24$ between the random slopes; participants who exhibited a stronger effect of alternation also exhibited a stronger effect of RU. Third, all the best fitting models (9, 11, and 12) have V, V/TU, or both. These components are the *random* sampling components. In sum, overall, there are clearly both aspects of directed and random sampling, there is just some remaining ambiguity with regards to V vs. V/TU.

To summarize these findings in Fig. 5, we represented the influences on choice in two simplified ways. First, the alternation tendency is represented by a negative autocorrelation cycle. Second, the links from pain to expectancy to choice represent the V/TU and RU components of the computational model, which assume that participants use the pain experiences to develop representations of expectations of the two choices and representations of the relative and total uncertainty of these expectations.

3.3.4. Summary of theory

Fig. 6 presents a high-level summary of the findings of the study. Compared to Fig. 1 (Model C), all of the expected links were found (though note the caveat of the Expectancy ⟶ Choice link discussed below). Importantly, there are two cycles, one which captures a rational learning process during the conditioning phase (Choice ⟶ Shock ⟶ Pain ⟶ Expectancy ⟶ Choice), and one which captures the placebo effect feedback loop (Pain ⟷ Expectancy).

4. Discussion

The present study investigated how individuals come to expect that one treatment will be more effective than another and whether these expectancies drive their choice behaviour. Four novel findings emerged. First, the two sham treatments produced a placebo analgesic effect at test that varied as a product of the strength of the pain relief that each provided during conditioning, with the more effective treatment generating a larger placebo effect than the less effective treatment. Second, during both the conditioning and test phases, a recursive placebo effect loop was observed - participants formed expectations based on their pain experiences, and their expectations modulated their subsequent experience of pain. Third, the strength of the placebo effect during the test phase was predicted by individual differences in expectancies generated during the conditioning phase. Fourth, we found clear evidence that participants use both relative uncertainty (RU; embodied in the UCB algorithm) for choosing which treatment to test, and also have clear evidence for an alternation tendency. Both of these can be viewed as types of directed exploration.

Below we discuss the main implications of these findings for the relevant literatures.

4.1. The recursive association between pain and expectancy

We found clear evidence of placebo analgesia in our modified paradigm, with the optimal treatment capable of producing stronger placebo analgesia than the suboptimal treatment. Further, modelling confirmed an anticipated expectancy↔pain feedback loop: expectancy predicted subsequent pain on a trial-to-trial basis, and pain predicted subsequent expectancy. This is consistent with the placebo literature (Anchisi & Zanon, 2015; Büchel et al., 2014; Jepma et al., 2018; Wiech, 2016 - but see Au Yeung et al., 2014, and Colagiuri & Quinn, 2018, for exceptions), but previous research has not been designed to explore the association during conditioning, i.e. while learning is initially occurring, nor when participants have control over treatment selection. Current results therefore extend the literature by demonstrating that the expectancy ↔ pain feedback loop is an ongoing dynamic process already present during conditioning. As such, findings indicate that in cases when individuals make choices about biological-relevant outcomes, the expectancies driving those choices can actually interact with what they experience via the placebo effect. This marks an important extension of traditional exploit-explore designs. Specifically, no opportunity for the individual's expectancy to actually influence the outcomes experienced has existed in previous research.

As detailed in the introduction, the evidence for a placebo effect during conditioning ties in closely with standard accounts of reinforcement learning. If the perceived outcome (pain) is a linear weighting of the expectancy and the objective outcome (shock), this effect can be modelled as simply a slower learning rate, compared to if the outcome is experienced objectively without any bias from expectancy. However, peering under the hood reveals a secondary recursive placebo process of expectancies influencing the perception of pain. This means that when trying to learn about and maximize benefits and minimize harms in a biologically-relevant explore-exploit setting, it is not just the objective outcomes that matter, but also how the outcomes are perceived. This is important for placebo researchers as well as cognitive scientists more generally. Not only do our results outline the algorithmic processes engaged during decision making process for pain, they also demonstrate that choice-related expectancies generated during decision making can directly modulate pain perception and autonomic responsivity. Given that expectancies generated as a product of treatment choice are likely to form naturally during the treatment process, this study provides first evidence that these choice-elicited expectancies are directly relevant to the generation of the placebo effect. Furthermore, these results demonstrate that this placebo effect develops each time an individual makes a treatment choice (expectancy↔pain feedback loop).

4.2. The role of expectancy and choice during conditioning on the placebo effect at test

We also sought to determine how expectations surrounding pain and choice behaviour during conditioning influence the placebo effect at test (i.e. when participants no longer had choice over their treatment). Results demonstrate that individual differences in the placebo effect during conditioning (i.e. the strength with which expectancy predicts pain) is associated with the magnitude of the placebo effect at test. To the best of our knowledge, the recursive association between expectancy and pain has not been tested during the conditioning phase before. In the current study, this finding is extended further to demonstrate that expectancy-based modulation of pain outcomes that develop during the learning process (i.e. when treatments are effective) can directly impact the placebo effect when the treatment no longer contains any 'active' analgesic property. This effect was only apparent when the overall placebo effect was considered (treatment vs. no treatment), rather than the differential effect (optimal vs. suboptimal treatment); however, in this study the difference in pain between treatment vs. no-treatment was much larger than between the two treatments. Consequently, null results may be driven by this smaller difference.

Only expectancies, and not choice behaviour (i.e. switch rate and optimal choice), predicted the placebo effect at test. This is consistent with the predicted theoretical model where overall expectancies generated throughout the course of the experiment, rather than specific patterns of choice behaviour, are associated with the placebo effect. However, levels of exploitation were low and it cannot be ruled out that the limited strength of these variables may have obscured any association with the placebo effect that might exist.

4.3. Exploratory behaviour and choice for biologically-relevant outcomes

Previous research employing artificial outcomes has reported that people use both directed and random exploration (Gershman, 2018, 2019). In those studies, individuals tend to test options that they have tested fewer times in the past and therefore have higher relative uncertainty about (RU), which is predicted by the UCB algorithm, a form of directed exploration. The current study also found this effect. The prior studies also found evidence for V/TU (Thompson sampling, a sophisticated type of random sampling), and sometimes also for V (pure random sampling). In our study, the best-fitting model included V/TU and not V; however, models with just V or with both fit nearly as well (Gershman, 2018, 2019). In summary, though there is some uncertainty about V vs. V/TU, the current results clearly agree with the prior studies in that people use RU, and indicates that similar exploratory behaviors are engaged both when artificial outcomes are modelled as well as real world scenarios with higher stakes outcomes.

Additionally, participants exhibited a strong tendency to alternate (generating the largest effect size in the model). We posit that alternation can be viewed as a very simple directed exploration strategy. Alternation reduces relative uncertainty (and to some extent total uncertainty). In real-world settings, particularly where fluctuations in background symptomology can result in autocorrelation in treatment outcomes over time, alternation is a valuable strategy for other reasons; rapid switching allows more successful decoupling of treatment efficacy from background fluctuations in pain (Rottman, 2016), meaning that the best treatment can be selected in the long run. In fact, during conditioning, pain ratings changed (demonstrated by positive autocorrelation in pain regressions), even though the average shocks for each option remained stable. In situations with positive autocorrelation of baseline symptomology alternation is especially beneficial for discriminating between treatments (Rottman, 2016).

Compared to research with artificial outcomes, a high switch rate of this type is unusual (Rottman, 2016). This could be for several reasons. First, the propensity to switch may be accentuated when treatment outcomes are biologically-relevant. To the extent that alternation is useful, this could reflect a stronger motivation to achieve the desirable outcome (less pain). Another possibility is that fine gradations of pain could be more difficult to discriminate than clear numerical outcomes (e.g. points), which could encourage alternation to help them compare experiences. Third, there was no 'horizon' in the current study; no information regarding trial numbers and when the study would end was provided. The horizon was omitted to correspond to a class of 'infinite-horizon' real world situations (e.g., unknown duration of a disease process). When a horizon exists, the optimal strategy is to exploit leading up to the end of testing (Lee et al., 2011; Steyvers et al., 2009; Wilson et al., 2014). The lack of a horizon could therefore have encouraged exploration of options, given that termination of the experiment could not be predicted, leading to the high alternation rate.

4.4. Limitations

There is one primary limitation and future direction. In the theoretical model there are links from shock ⟶ pain ⟶ expectancy ⟶ choice and we did test the influence of shock on pain, and pain on expectancy. However, when modelling choice, the *V* measure (difference in outcomes for one choice vs. another) was calculated from the shocks not the rated expectancies. In fact, the research focusing on choice in explore-exploit tasks that we built upon did not directly investigate how expectancies are developed, though there are separate lines of research on this question (e.g. Bornstein et al., 2017; Rottman, 2016; Worthy & Todd Maddox, 2014). Ideally to study this question participants would rate their expectancies for each option after every experience, or, at least rate their expectancy for the chosen option after experiencing it. In line with placebo research, in our study participants rated their expectancy only for the chosen option before experiencing it, which makes it easy to study the expectancy ⟶ pain link but difficult to study the pain ⟶ expectancy link, and we did not attempt to study the expectancy ⟶ choice link directly because it would have required imputing expectancies at periods of time that they were not directly measured. In sum, if future research measures expectancy for all choices after each outcome, it would permit a stronger test of these links.

4.5. Summary

In summary, the results are consistent with prior findings that participants use both directed and random sampling in explore-exploit situations, and extends those findings to physiologically-relevant situations. In line with prior research, these findings can be taken as evidence of a rational learning process. However, consistent with research on the placebo effect, we also found evidence of a secondary process in which experiences of pain were influenced by prior expectations that developed during the explore-exploit process. Further, results demonstrate that this expectancy-based modulation of pain during the learning process can impact the placebo effect at test. This extends the placebo literature, demonstrating that the recursive association between pain and expectancy is not isolated to the test phase, but also the explore-exploit literature, demonstrating that directed exploration occurs even when expectancies influence the outcome experienced.

Acknowledgments

This research was supported by Australian Research Council (DE160100864, DP180102061) and The University of Sydney (Psychology Visiting International Collaborator Scheme Grant).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2021.104753.

References

Addicott, M. A., Pearson, J. M., Sweitzer, M. M., Barack, D. L., & Platt, M. L. (2017). A primer on foraging and the explore/exploit trade-off for psychiatry research. *Neuropsychopharmacology, 42*(10), 1931–1939. https://doi.org/10.1038/npp.2017.108.

Anchisi, D., & Zanon, M. (2015). A Bayesian perspective on sensory and cognitive integration in pain perception and placebo analgesia. *PLoS One, 10*(2), Article e0117270. https://doi.org/10.1371/journal.pone.0117270.

Au Yeung, S. T., Colagiuri, B., Lovibond, P. F., & Colloca, L. (2014). Partial reinforcement, extinction, and placebo analgesia. *PAIN®, 155*(6), 1110–1117. https://doi.org/10.1016/j.pain.2014.02.022.

Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine Learning, 47*(2), 235–256. https://doi.org/10.1023/A:1013689704352.

Bąbel, P., Bajcar, E. A., Adamczyk, W., Kicman, P., Lisińska, N., Świder, K., & Colloca, L. (2017). Classical conditioning without verbal suggestions elicits placebo analgesia and nocebo hyperalgesia. *PLoS One, 12*(7), Article e0181856. https://doi.org/10.1371/journal.pone.0181856.

Barnes, P. M., Bloom, B., Nahin, R. L., & Stussman, B. J. (2009). *Costs of complementary and alternative medicine (CAM) and frequency of visits to CAM practitioners, United States. 2007* p. 18). Retrieved from https://stacks.cdc.gov/view/cdc/11548.

Bartley, H., Faasse, K., Horne, R., & Petrie, K. J. (2016). You can't always get what you want: The influence of choice on nocebo and placebo responding. *Annals of Behavioral Medicine, 50*(3), 445–451. https://doi.org/10.1007/s12160-016-9772-1.

Benedetti, F. (2008). Mechanisms of placebo and placebo-related effects across diseases and treatments. *Annual Review of Pharmacology and Toxicology, 48*(1), 33–60. https://doi.org/10.1146/annurev.pharmtox.48.113006.094711.

Benedetti, F., Amanzio, M., Baldi, S., Casadio, C., & Maggi, G. (1999). Inducing placebo respiratory depressant responses in humans via opioid receptors. *European Journal of Neuroscience, 11*(2), 625–631. https://doi.org/10.1046/j.1460-9568.1999.00465.x.

Benedetti, F., Maggi, G., Lopiano, L., Lanotte, M., Rainero, I., Vighetti, S., & Pollo, A. (2003). Open versus hidden medical treatments: The patient's knowledge about a therapy affects the therapy outcome. *Prevention & Treatment, 6*(1). https://doi.org/10.1037/1522-3736.6.1.61a. No Pagination Specified.

Biele, G., Erev, I., & Ert, E. (2009). Learning, risk attitude and hot stoves in restless bandit problems. *Journal of Mathematical Psychology, 53*(3), 155–167. https://doi.org/10.1016/j.jmp.2008.05.006.

Blanco, F., Matute, H., & Vadillo, M. A. (2011). Making the uncontrollable seem controllable: The role of action in the illusion of control. *Quarterly Journal of Experimental Psychology, 64*(7), 1290–1304. https://doi.org/10.1080/17470218.2011.552727.

Bornstein, A. M., Khaw, M. W., Shohamy, D., & Daw, N. D. (2017). Reminders of past choices bias decisions for reward in humans. *Nature Communications, 8*(1), 15958. https://doi.org/10.1038/ncomms15958.

Bott, F. M., & Meiser, T. (2020). Pseudocontingency inference and choice: The role of information sampling. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* https://doi.org/10.1037/xlm0000840. No Pagination Specified-No Pagination Specified.

Brown, J. A., Fowler, S. L., Rasinski, H. M., Rose, J. P., & Geers, A. L. (2013). Choice as a moderator of placebo expectation effects: Additional support from two experiments. *Basic and Applied Social Psychology, 35*(5), 436–444. https://doi.org/10.1080/01973533.2013.803968.

Büchel, C., Geuter, S., Sprenger, C., & Eippert, F. (2014). Placebo analgesia: A predictive coding perspective. *Neuron, 81*(6), 1223–1239. https://doi.org/10.1016/j.neuron.2014.02.042.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33*(2), 261–304. https://doi.org/10.1177/0049124104268644.

Colagiuri, B., & Quinn, V. F. (2018). Autonomic arousal as a mechanism of the persistence of nocebo hyperalgesia. *The Journal of Pain, 19*(5), 476–486. https://doi.org/10.1016/j.jpain.2017.12.006.

Colagiuri, B., Quinn, V. F., & Colloca, L. (2015). Nocebo hyperalgesia, partial reinforcement, and extinction. *The Journal of Pain, 16*(10), 995–1004. https://doi.org/10.1016/j.jpain.2015.06.012.

Colagiuri, B., Schenk, L. A., Kessler, M. D., Dorsey, S. G., & Colloca, L. (2015). The placebo effect: From concepts to genes. *Neuroscience, 307*, 171–190. https://doi.org/10.1016/j.neuroscience.2015.08.017.

Colloca, L., & Benedetti, F. (2009). Placebo analgesia induced by social observational learning. *Pain, 144*(1–2), 28–34. https://doi.org/10.1016/j.pain.2009.01.033.

Colloca, L., Petrovic, P., Wager, T. D., Ingvar, M., & Benedetti, F. (2010). How the number of learning trials affects placebo and nocebo responses. *Pain, 151*(2), 430–439. https://doi.org/10.1016/j.pain.2010.08.007.

Colloca, L., Sigaudo, M., & Benedetti, F. (2008). The role of learning in nocebo and placebo effects. *Pain, 136*(1–2), 211–218. https://doi.org/10.1016/j.pain.2008.02.006.

Cutler, D. M., & Everett, W. (2010). Thinking outside the pillbox — Medication adherence as a priority for health care reform. *New England Journal of Medicine, 362*(17), 1553–1555. https://doi.org/10.1056/NEJMp1002305.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature, 441*(7095), 876–879. https://doi.org/10.1038/nature04766.

Erev, I. (2010). A choice prediction competition for market entry games: An introduction. *Games, 1*(2), 117. https://doi.org/10.3390/g1020117.

Finniss, D. G., Kaptchuk, T. J., Miller, F. G., & Benedetti, F. (2010). Biological, clinical, and ethical advances of placebo effects. *The Lancet, 375*(9715), 686–695. https://doi.org/10.1016/S0140-6736(09)61706-2.

Frank, M. J., Doll, B. B., Oas-Terpstra, J., & Moreno, F. (2009). Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nature Neuroscience, 12*(8), 1062.

Geers, A. L., Rose, J. P., Fowler, S. L., & Brown, J. A. (2014). Patient involvement in treatment decision making can help or hinder placebo analgesia. *Zeitschrift für Psychologie, 222*(3), 165–170. https://doi.org/10.1027/2151-2604/a000179.

Geers, A. L., Rose, J. P., Fowler, S. L., Rasinski, H. M., Brown, J. A., & Helfer, S. G. (2013). Why does choice enhance treatment effectiveness? Using placebo treatments to demonstrate the role of personal control. *Journal of Personality and Social Psychology, 105*(4), 549–566. https://doi.org/10.1037/a0034005.

Gershman, S. J. (2018). Deconstructing the human algorithms for exploration. *Cognition, 173*, 34–42. https://doi.org/10.1016/j.cognition.2017.12.014.

Gershman, S. J. (2019). Uncertainty and exploration. *Decision, 6*(3), 277–286. https://doi.org/10.1037/dec0000101.

Gershman, S. J. (2020). Origin of perseveration in the trade-off between reward and complexity. *bioRxiv*. https://doi.org/10.1101/2020.01.16.903476, 2020.2001.2016.903476.

Gershman, S. J., & Tzovaras, B. G. (2018). Dopaminergic genes are associated with both directed and random exploration. *Neuropsychologia, 120*, 97–104. https://doi.org/10.1016/j.neuropsychologia.2018.10.009.

Geuter, S., Eippert, F., Hindi Attar, C., & Büchel, C. (2013). Cortical and subcortical responses to high and low effective placebo treatments. *NeuroImage, 67*, 227–236. https://doi.org/10.1016/j.neuroimage.2012.11.029.

Geuter, S., Koban, L., & Wager, T. D. (2017). The cognitive neuroscience of placebo effects: Concepts, predictions, and physiology. *Annual Review of Neuroscience, 40*(1), 167–188. https://doi.org/10.1146/annurev-neuro-072116-031132.

Gittins, J. C. (1979). Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society: Series B: Methodological, 41*(2), 148–164. https://doi.org/10.1111/j.2517-6161.1979.tb01068.x.

Jepma, M., Koban, L., van Doorn, J., Jones, M., & Wager, T. D. (2018). Behavioural and neural evidence for self-reinforcing expectancy effects on pain. *Nature Human Behaviour, 2*(11), 838–855. https://doi.org/10.1038/s41562-018-0455-8.

Jepma, M., & Wager, T. D. (2015). Conceptual conditioning: Mechanisms mediating conditioning effects on pain. *Psychological Science, 26*(11), 1728–1739. https://doi.org/10.1177/0956797615597658.

Johnson, P. C. D. (2014). Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in Ecology and Evolution, 5*(9), 944–946. https://doi.org/10.1111/2041-210x.12225.

Lee, M. D., Zhang, S., Munro, M., & Steyvers, M. (2011). Psychological models of human and optimal performance in bandit problems. *Cognitive Systems Research, 12*(2), 164–174. https://doi.org/10.1016/j.cogsys.2010.07.007.

Luhmann, C. C., & Ahn, W. K. (2007). BUCKLE: A model of unobserved cause learning. *Psychological Review, 114*(3), 657–677. https://doi.org/10.1037/0033-295X.114.3.657.

Luhmann, C. C., & Ahn, W. K. (2011). Expectations and interpretations during causal learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(3), 568–587. https://doi.org/10.1037/a0022970.

Marsh, J. K. (2009). Spontaneous assimilation of continuous values and temporal information in causal induction. *Journal of Experimental Psychology, 35*(2), 334. https://doi.org/10.1037/a0014929.

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., … Gonzalez, C. (2015). Unpacking the exploration–exploitation tradeoff: A synthesis of human and animal literatures. *Decision, 2*(3), 191–215. https://doi.org/10.1037/dec0000033.

Montgomery, G. H., & Kirsch, I. (1997). Classical conditioning and the placebo effect. *Pain, 72*(1–2), 107–113. https://doi.org/10.1016/S0304-3959(97)00016-X.

Navarro, D. J., Newell, B. R., & Schulze, C. (2016). Learning and choosing in an uncertain world: An investigation of the explore–exploit dilemma in static and dynamic environments. *Cognitive Psychology, 85*, 43–77. https://doi.org/10.1016/j.cogpsych.2016.01.001.

Osterberg, L., & Blaschke, T. (2005). Adherence to medication. *New England Journal of Medicine, 353*(5), 487–497. https://doi.org/10.1056/NEJMra050100.

Price, D. D., Finniss, D. G., & Benedetti, F. (2008). A comprehensive review of the placebo effect: Recent advances and current thought. *Annual Review of Psychology, 59*(1), 565–590. https://doi.org/10.1146/annurev.psych.59.113006.095941.

Raisch, S., Birkinshaw, J., Probst, G., & Tushman, M. L. (2009). Organizational ambidexterity: Balancing exploitation and exploration for sustained performance. *Organization Science, 20*(4), 685–695. https://doi.org/10.1287/orsc.1090.0428.

Rescorla, R. A. (1972). *A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical conditioning II: Current research and theory. 2* pp. 64–99).

Rose, J. P., Geers, A. L., Fowler, S. L., & Rasinski, H. M. (2013). Choice-making, expectations, and treatment positivity: How and when choosing shapes aversive experiences. *Journal of Behavioral Decision Making, 27*(1), 1–10. https://doi.org/10.1002/bdm.1775.

Rose, J. P., Geers, A. L., Rasinski, H. M., & Fowler, S. L. (2012). Choice and placebo expectation effects in the context of pain analgesia. *Journal of Behavioral Medicine, 35*(4), 462–470. https://doi.org/10.1007/s10865-011-9374-0.

Rottman, B. M. (2016). Searching for the best cause: Roles of mechanism beliefs, autocorrelation, and exploitation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(8), 1233–1256. https://doi.org/10.1037/xlm0000244.

Rottman, B. M., Marcum, Z. A., Thorpe, C. T., & Gellad, W. F. (2017). Medication adherence as a learning process: Insights from cognitive psychology. *Health*

Psychology Review, 11(1), 17–32. https://doi.org/10.1080/17437199.2016.1240624.

Roy, M., Shohamy, D., Daw, N., Jepma, M., Wimmer, G. E., & Wager, T. D. (2014). Representation of aversive prediction errors in the human periaqueductal gray. *Nature Neuroscience, 17*(11), 1607–1612. https://doi.org/10.1038/nn.3832.

Russo, D. J., Roy, B. V., Kazerouni, A., Osband, I., & Wen, Z. (2018). A tutorial on Thompson sampling. *Foundations and Trends in Maching Learning, 11*(1), 1–96. https://doi.org/10.1561/2200000070.

Seymour, B., Daw, N. D., Roiser, J. P., Dayan, P., & Dolan, R. (2012). Serotonin selectively modulates reward value in human decision-making. *The Journal of Neuroscience, 32*(17), 5833–5842. https://doi.org/10.1523/jneurosci.0053-12.2012.

Smith, W. K., & Tushman, M. L. (2005). Managing strategic contradictions: A top management model for managing innovation streams. *Organization Science, 16*(5), 522–536. https://doi.org/10.1287/orsc.1050.0134.

Speekenbrink, M., & Konstantinidis, E. (2015). Uncertainty and exploration in a restless bandit problem. *Topics in Cognitive Science, 7*(2), 351–367.

Stewart-Williams, S., & Podd, J. (2004). The placebo effect: Dissolving the expectancy versus conditioning debate. *Psychological Bulletin, 130*(2), 324–340. https://doi.org/10.1037/0033-2909.130.2.324.

Steyvers, M., Lee, M. D., & Wagenmakers, E.-J. (2009). A Bayesian analysis of human decision-making on bandit problems. *Journal of Mathematical Psychology, 53*(3), 168–179. https://doi.org/10.1016/j.jmp.2008.11.002.

Tang, B., Geers, A., Barnes, K., & Colagiuri, B. (2019). Instrumental control enhances placebo analgesia. *The Journal of Pain*. https://doi.org/10.1016/j.jpain.2019.05.013.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika, 25*(3/4), 285–294. https://doi.org/10.2307/2332286.

Tracey, I. (2010). Getting the pain you expect: Mechanisms of placebo, nocebo and reappraisal effects in humans. *Nature Medicine, 16*(11), 1277–1283. https://doi.org/10.1038/nm.2229.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453. https://doi.org/10.1126/science.7455683.

Voudouris, N. J., Peck, C. L., & Coleman, G. (1985). Conditioned placebo responses. *Journal of Personality and Social Psychology, 48*(1), 47–53. https://doi.org/10.1037/0022-3514.48.1.47.

Wang, O., Lee, S. W., O'Doherty, J., Seymour, B., & Yoshida, W. (2018). Model-based and model-free pain avoidance learning. *Brain and Neuroscience Advances, 2*. https://doi.org/10.1177/2398212818772964, 2398212818772964.

Wiech, K. (2016). Deconstructing the sensation of pain: The influence of cognitive processes on pain perception. *Science, 354*(6312), 584–587. https://doi.org/10.1126/science.aaf8934.

Wiech, K., Ploner, M., & Tracey, I. (2008). Neurocognitive aspects of pain perception. *Trends in Cognitive Sciences, 12*(8), 306–313. https://doi.org/10.1016/j.tics.2008.05.005.

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore-exploit dilemma. *Journal of Experimental Psychology. General, 143*(6), 2074–2081. https://doi.org/10.1037/a0038199.

Wimmer, G. E., & Büchel, C. (2020). Reactivation of pain-related patterns in the hippocampus from single past episodes relates to successful memory-based decision making. *bioRxiv*. https://doi.org/10.1101/2020.05.29.123893, 2020.2005.2029.123893.

Worthy, D. A., & Todd Maddox, W. (2014). A comparison model of reinforcement-learning and win-stay-lose-shift decision-making processes: A tribute to W.K. Estes. *Journal of Mathematical Psychology, 59*, 41–49. https://doi.org/10.1016/j.jmp.2013.10.001.

Yi, S. K. M., Steyvers, M., & Lee, M. (2009). Modeling human performance in restless bandits with particle filters. *The Journal of Problem Solving, 2*(2), 33–53.