**UNIVERSITÄTSKLINIKUM HAMBURG-EPPENDORF**

Zentrum für Experimentelle Medizin
Institut für Systemische Neurowissenschaften

Direktor: Prof. Dr. med. Christian Büchel

# Exploring the Role of Predictive Coding and Active Inference in Pain: A Bayesian Perspective

**Kumulative Dissertation**

zur Erlangung des Grades eines Doctor of Philosophy (PhD)
an der Medizinischen Fakultät der Universität Hamburg.

vorgelegt von:

Andreas Strube
aus Hamburg, Deutschland

Hamburg 2022

Angenommen von der Medizinischen Fakultät am: **18.08.2023**

Veröffentlicht mit Genehmigung der Medizinischen Fakultät der Universität Hamburg

| | |
|---|---|
| Prüfungsausschuss, der/die Vorsitzende: | **Prof. Dr. Christian Büchel** |
| Prüfungsausschuss, 2. Gutachter/in: | **Prof. Dr. Jürgen Lorenz** |
| Drittgutachter/in: | **Prof. Dr. Ulrike Bingel** |

## Inhaltsverzeichnis / Table of Contents

## 1.0 Introduction

Pain is a complex sensation that can arise from various forms of noxious (i.e. potentially tissue damaging) stimulation, such as heat or pressure on the skin. However, large interindividual variation in pain sensitivity suggests that the relationship between nociception and pain is not straightforward. Importantly, pain is frequently felt in the absence of a nociceptive stimulus, as seen in chronic pain conditions like facial pain syndrome, chronic migraine, and phantom pain, where patients experience pain without any apparent cause (May, 2008, 2011; Parkes, 1973; Ramachandran & Rogers-Ramachandran, 1996; Simmel, 1959). The most impressive instance is possibly phantom pain, that is, a sensation of pain in an amputated limb despite the absence of the limb itself (Halligan, 2002; Ramachandran & Rogers-Ramachandran, 1996).

Pain is not simply a direct response to a noxious stimulus, but is instead influenced by various contextual factors, such as the individual's perceived control over the pain and their expectations of it (Atlas & Wager, 2014; Bingel et al., 2006; Büchel et al., 2014; Colloca & Benedetti, 2005; Helmchen et al., 2006; Karsh et al., 2018; Mohr et al., 2008, 2012; Pervin, 1963; Petrovic et al., 2002; Staub et al., 1971; Thompson, 1981; Wager et al., 2004; Wang et al., 2011; Weisenberg et al., 1985). Even attentional processes and distraction can affect the perception of pain (Bantick et al., 2002; Hauck et al., 2007, 2013; Miron et al., 1989; Sprenger et al., 2012). While the objective stimulus quality is transmitted to the cortex via the ascending pain system, contextual factors can modulate these afferent signals through descending pathways. For example, pain expectations have been shown to alter responses at the spinal dorsal horn in a top-down fashion (Eippert et al., 2009; Sprenger et al., 2015; Tinnermann et al., 2017). These findings demonstrate the complex nature of pain and the need to consider these contextual factors in pain management. To understand pain more comprehensively, it is necessary to consider both the sensory processes that transmit information about the noxious stimulus to the brain, as well as the contextual factors that can modulate these signals and the integration of these two components.

A Bayesian perspective on pain allows for this integration of the various factors that influence the perception of pain (Anchisi & Zanon, 2015; Büchel et al., 2014; Ongaro & Kaptchuk, 2019; Wiech, 2016). A *Bayesian Pain Model* views the pain experience as a combination of prior information and sensory input, with both factors weighted by

their precision to create a percept. Pain expectations, which can be changed through learning, are an important part of this model. One notable application of this model is placebo hypoalgesia, in which positive prior pain experiences can lead to a decrease in pain through the placebo effect (Atlas & Wager, 2014; Bingel et al., 2006; Colloca & Benedetti, 2005; Petrovic et al., 2002; Wager et al., 2004). This phenomenon has been observed in studies that show that positive treatment expectations can reduce pain even in the absence of active treatment.

The *Bayesian Pain Model* is based on the idea that the brain uses statistical methods to construct hypotheses about the world around us. The model incorporates concepts from the *Bayesian Brain* (Dayan et al., 1995; Friston, 2012; Helmholtz, 1867; Knill & Pouget, 2004), *Predictive Coding* (Rao & Ballard, 1999; Srinivasan et al., 1982), the *Free Energy Principle* (Friston, 2010; Friston et al., 2006) and *Active Inference* (Brown et al., 2013; Friston, 2010; Friston et al., 2009). These ideas can be implemented in computational models of canonical cortical microcircuits that encode expectations and the violations of those expectations, also known as prediction errors (Arnal & Giraud, 2012; Bastos et al., 2012). The *Bayesian Pain Model* provides a framework for understanding how the brain processes sensory information to create our experience of pain.

In this cumulative dissertation, we address the complex nature of pain and its modulation by contextual factors through a Bayesian lens.

The core hypotheses of this work are that…

1) the brain processes nociceptive stimuli via *Predictive Coding* mechanisms, utilizing expectations and prediction errors.
2) the brain generates a pain percept in a Bayes-optimal manner as explained by a *Bayesian Pain Model*.
3) these mechanisms can explain contextual modulations of pain, e.g. through agency or placebo expectations.

In our first paper, we examined the relationship between expectations and prediction errors in phasic thermal pain (Reprint 15.1; Strube et al., 2021a). Drawing on *Predictive Coding* theory, we explored whether these components are reflected in specific patterns of neural activity, as measured by electroencephalography (EEG). We hypothesized that expectations would be represented in lower frequency bands (such

as alpha or beta, between 8-30Hz), while prediction errors would be encoded in higher frequencies (such as gamma, >30Hz), based on the asymmetries of oscillatory power spectra (Arnal & Giraud, 2012; Bastos et al., 2012). Furthermore, we hypothesized that the neural signals associated with expectations would precede those related to prediction errors.

We were able to uncover the temporal and spectral patterns of stimulus intensity, expectations, and prediction errors in pain anticipation and painful thermal stimulation: We found that an expectation signal was generated in the alpha-to-beta range (8-30Hz) followed by a prediction error signal in the gamma range (>30Hz) during painful stimulation. Contrary to our predictions, we observed a decrease in gamma activity (>30Hz) associated with prediction errors, when the predictive cue did not match the actual stimulus intensity. In contrast, increasing stimulus intensity was characterized by the typical spectral patterns associated with nociception, characterized by an increase in theta activity (4-8Hz), a decrease in alpha-to-beta activity (8-30Hz), and an increase in gamma activity (>30Hz) (see Ploner et al., 2017 for a review). Our findings provide insight into the temporal and spectral orchestration underlying *Predictive Coding* in pain perception (Strube et al., 2021a).

In our second paper, we extended our investigation to include an affective dimension by using aversive images (Reprint 15.2; Strube et al., 2021b). Our findings revealed fundamentally different patterns of neural activity in affective visual processing, suggesting the existence of modality-specific oscillatory networks in *Predictive Coding* for pain and affective visual stimulation. In contrast to our observations for painful stimuli, we found that all components of *Predictive Coding* (including stimulus intensity, expectations, and prediction errors) were encoded in low frequency bands ranging from theta (4-8Hz) to low gamma (30-50Hz) during the presentation of aversive pictures. This indicates that these signals are specific to the modality being processed, and that the temporal and spectral orchestration of *Predictive Coding* differs between pain and affective visual processing.

In our third paper, we aimed to explore how the contextual aspect of agency can be integrated within the framework of a *Bayesian Pain Model*, and whether this supports hypotheses derived from the *Active Inference* model (Reprint 15.3; Strube et al., 2022, submitted). Agency describes the extent to which an actor believes to be effective in producing an outcome via his own actions, including sensory outcomes or events.

We generated hypotheses based on the classical *Forward Model* (Blakemore et al., 1998, 2000) and tested them against those derived from *Active Inference* (Brown et al., 2013). To test these hypotheses, we conducted a placebo/nocebo experiment in which treatment was self- or externally-initiated. Our results suggest that agency can modulate the effectivity of treatment. Computational Bayesian modeling using variational Bayesian analysis (VBA; Daunizeau et al., 2014) showed that this effect can be explained as a shift in intensity expectations, in line with predictions from the *Forward Model*. A modulation of prior precision, as derived from *Active Inference*, was not able to explain this effect.

Interestingly, we found the reverse pattern of typical stimulus intensity representations in the time-frequency patterns of nociceptive phasic stimulation, as measured by EEG. In general, nociceptive phasic stimulation is associated with a decrease in alpha-to-beta (8-30Hz) oscillations and an increase in theta (4-8Hz) oscillations. In this study, objectively better treatment outcomes were associated with an increase in alpha-to-beta (8-30Hz) oscillations and a decrease in theta (4-8Hz) oscillations. Additionally, theta-to-alpha (4-12Hz) activity, which was temporally associated with an expectation-generating cue, was predictive of Visual Analogue Scale (VAS) ratings when stimulus intensity was held constant. This indicates a shift in the prior representing expectations explaining the reduction in pain by self-initiation of pain treatment, as compared to a shift in the likelihood representing sensory information. Overall, our data provide insight into the neural mechanisms underlying the placebo/nocebo effect and the role of agency in pain perception.

With this study, we not only demonstrated that contextual factors such as agency can be integrated into the *Bayesian Pain Model*, but we also showed how these factors can modulate the typical components of Bayesian integration. This allows for a more sophisticated understanding of how pain is perceived and processed in the brain. By considering the influence of contextual factors on Bayesian integration, we can gain a deeper insight into the mechanisms underlying pain perception.

Through the conduct of three studies, we sought to examine the role *of Predictive Coding* and *Active Inference* in the modulation of pain through the lens of Bayesian inference. The findings of these studies were then integrated to provide an understanding of the influence of these mechanisms on the experience of pain in a *Bayesian Pain Model*. This dissertation provides a cumulative assessment of our

results and offers conclusions on the role of *Predictive Coding* and *Active Inference* in pain processing.

## 2.0 Pain versus Nociception: The Influence of Contextual Factors on Pain

At first glance, the basic function of pain seems quite simple – to signal if a physical stimulus reaches harmful intensities, and facilitate countermeasures (Torebjörk, 1985; Witt & Griffin, 1962). Aspects serving this function can happen automatically and in very brief time frames. For example, when we grab onto a hot stove top, there is typically an immediate withdrawal reaction, termed the nociceptive withdrawal reflex (Bromm & Treede, 1980; Fields & Heinricher, 1989; Neziri et al., 2010; Schouenborg et al., 1992).

In a simple mechanistic system, a signal indicating injury or destruction of the skin could be integrated in a model in which higher temperatures (or, generally, physical intensities) are directly associated with higher pain via a monotonic temperature-pain-relationship. Experimentally, for example, it has been shown that there is a monotonically increasing relation between stimulus temperature and the magnitude of pain sensations in a range from 40-50°C (LaMotte & Campbell, 1978). Pain, however, is more complex, as contextual factors, such as attention or expectations modulate our pain experience; this even applies to pain thresholds, i.e. the temperature where an individual perceives the non-pain perception to flip to a pain perception (e.g. Taesler & Rose, 2016). As Tracey & Mantyh (2007) pointed out in their definition of pain: “A conscious experience, an interpretation of the nociceptive input influenced by memories, emotional, pathological, genetic, and cognitive factors” (Tracey & Mantyh, 2007, p.377).

In this chapter, I will delineate the neurophysiological basis underlying the sensation of pain. I will begin by describing the physiological basis of bottom-up processing of sensory pain information, including the ascending pain pathway involved in the transmission of pain signals from peripheral sensory neurons to the brain. I will then discuss the role of top-down processing in the modulation of pain, and how cognitive and emotional factors can influence our perception of pain via descending pain pathways.

## 2.1 The Ascending Pain Pathway: Transmitting Bottom-Up Sensory Information

Pain processing typically starts with the activation of nociceptors which are specialized peripheral sensory neurons in the skin that sense noxious extremes in temperature and pressure, as well as chemical stimulation (e.g. acids) (Bessou & Perl, 1969; Burgess & Perl, 1967; Witt & Griffin, 1962). Note that this is not always the case: Pain can be experienced without the involvement of peripheral sensory neurons, for example in the case of facial pain syndrome, chronic migraine, or phantom pain (May, 2008, 2011; Parkes, 1973; Ramachandran & Rogers-Ramachandran, 1996; Simmel, 1959).

The information from peripheral nociceptors is transduced into long-range electrical impulses that are sent to the first synapse in the dorsal horn of the spinal cord, then towards the brain and (mostly relayed via the thalamus) to a widely distributed network of cortical structures (Dubin & Patapoutian, 2010). Typically, this pathway is termed ascending pain pathway (Andersen & Dafny, 1983; Hammond, 1989; Qiao & Dafny, 1988; Willis et al., 1985). In a meta-analysis by Apkarian et al. (2005), cortical structures associated with pain processing were defined as a result of 68 analyzed studies. According to this meta-analysis, main components of the pain system are comprised of somatosensory, insular, cingulate, and prefrontal cortices, the thalamus, subcortical areas, and the brainstem (Apkarian et al., 2005).

There are two main types of nociceptors, namely A-fiber and C-fiber nociceptors, describing the nerve fibers involved in relaying the nociceptive signal. Initial fast-onset pain by high temperatures (e.g. 47°C) is typically signaled by myelinated A-fiber nociceptors which have relatively high conduction velocities (5–30 m/s), thus enabling fast motor reactions to harmful stimuli. These fast-conducting fibers are responsible for the initial sharp pain that is often experienced when we encounter harmful stimuli. C-fibers, on the other hand, are unmyelinated and have slower conduction velocities (0.4-1.4 m/s), but can be activated at lower temperatures (approx. 42°C) (Dubin & Patapoutian, 2010).

In summary, A-fibers and C-fibers transfer nociceptive signals from the periphery to the dorsal horn of the spinal cord, where they are processed and relayed to the brain. This bottom-up processing of sensory stimuli is the foundation of the ascending pain pathway. However, the experience of pain is not solely determined by these bottom-up signals. Rather, it is also influenced by top-down processes that modulate the

perception and experience of pain. The descending pain pathway is the basis of these top-down processes, and it plays a crucial role in the contextual modulation of pain. In the following section, I will describe the neurophysiological basis of this pathway and its role in the complex experience of pain.

### 2.2 The Descending Pain Pathway: Top-Down Pain Regulation

The experience of pain is influenced not only by sensory information transmitted through the ascending pain pathway, but also by contextual factors that shape our perception of pain. This is where the descending pain network comes into play: Descending pathways can inhibit or facilitate nociceptive processing, and have been shown to involve specific brain regions such as the dorsolateral prefrontal cortex, the anterior cingulate cortex, and the periaqueductal gray, which, in consequence via the rostral ventromedial medulla, lead to modulation of activity at the dorsal horn (Eippert et al., 2009; Geuter & Büchel, 2013; Sprenger et al., 2012; Tinnermann et al., 2017). In other words, the descending pain network allows for the integration of top-down cognitive and emotional factors in the relaying of nociceptive information, and ultimately in the perception and experience of pain.

One prominent example of the influence of top-down processes on pain perception is the phenomenon of placebo and nocebo effects. Placebo hypoalgesia refers to the analgesic effects of positive pain expectations and positive treatment expectations, whereas nocebo hyperalgesia describes the facilitation of pain by negative pain expectations and negative treatment expectations (Atlas and Wager, 2014; Bingel et al., 2006; Colloca & Benedetti, 2005; Petrovic et al., 2002; Wager et al., 2004). These effects have been shown to involve the descending pain network and the modulation of nociceptive processing in the dorsal horn (Eippert et al., 2009; Geuter & Buchel, 2013; Tinnermann et al., 2017).

Direct evidence for spinal cord involvement in placebo hypoalgesia has been shown in a combined functional magnetic resonance imaging (fMRI) study, where both cortical and spinal responses to placebo hypoalgesia have been measured (Sprenger et al., 2015). In another study, interactions between the brain and the spinal cord have been shown to mediate value effects in nocebo hyperalgesia, involving the prefrontal cortex, brainstem, and spinal cord (Tinnermann et al., 2017). Another example of the involvement of the descending pain pathway is the modulation via top-down attentional processes. In a study by Sprenger et al. (2012), fMRI data was acquired from the spinal

cord in combination with thermal pain stimulation while participants were involved in a well-established working memory task to modulate attentional processes. Here, the authors showed that attentional processes modulate pain perception by inhibition of incoming pain signals in the spinal cord (Sprenger et al., 2012).

In summary, information from nociceptive stimuli at peripheral nociceptors is carried in a bottom-up fashion to higher cortical regions via ascending pathways. At the same time descending pathways modulate the pain experience and neural information transmission in a top-down fashion. A prominent example is the influence of expectations - here positive (placebo) and negative (nocebo) expectations modulate activity at pain areas including the dorsal horn via descending pathways, which is ultimately reflected in placebo hypoalgesia and nocebo hyperalgesia: a decreased or increased sensitivity to the nociceptive input, leading to a reduction (hypoalgesia; based on placebo effects) or increase of pain sensation (hyperalgesia; based on nocebo effects).

## 3.0 The Brain as a Statistical Machine: The Bayesian Brain

In this chapter, I will explore the concept of the *Bayesian Brain* and its relevance to the pain system, which consequently leads to a Bayesian perspective in pain processing in a *Bayesian Pain Model*. In this context, it is necessary to have a basic understanding of the brain as a statistical machine. This incorporates ideas from the *Free Energy Principle* (Friston, 2010; Friston et al., 2006) and *Predictive Coding* (Huang & Rao, 2011; Rao & Ballard, 1999; Srinivasan et al., 1982).

The concept of the *Bayesian Brain* addresses the question of how the brain processes uncertain or ambiguous information (Knill & Pouget, 2004). What is the role of uncertainty in action and perception? It suggests that perception can be thought of as a form of hypothesis testing, where the brain generates hypotheses about the properties of the world and uses sensory information to test and refine these hypotheses (Gregory, 1980; Gregory et al., 1968; Kersten et al., 2004). For example, the visual representation of an image can be described as a probabilistic integration of prior object knowledge (hypothesis) with image features (sensory information) (Kersten et al., 2004). The brain is theorized to integrate prior knowledge, or expectations, with sensory data in a probabilistic manner. Through this process of sampling and updating, the *Bayesian Brain* is able to form a representation of the world around us.

This process is thought to operate in a Bayes-optimal manner, which means that it integrates sensory information and expectations in the most efficient way possible to form a percept (see Schwartenbeck et al., 2015 for theoretical considerations of Bayes-optimal and sub-optimal inference). Bayes-optimal inference can be derived from Bayes' theorem, which is a fundamental principle of probability that describes the relationship between the prior probability of a hypothesis, the likelihood of observing certain evidence given that hypothesis, and the posterior probability of the hypothesis after taking the evidence into account. This is formalized in the following equation:

(Eq.1; Bayes' theorem):

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

Where P(H|E) is the posterior probability of the hypothesis H after observing evidence E, P(E|H) is the likelihood of observing evidence E given hypothesis H, P(H) is the prior probability of hypothesis H, and P(E) is the model evidence. This theorem provides a way to update our beliefs about the world based on new information and is a key component of Bayesian inference. By using Bayes' theorem, the *Bayesian Brain* can integrate prior expectations with sensory information in order to make predictions about the world.

In this way, the prior object knowledge is tested by perception, and evidence is accumulated in favor or against prior object knowledge, leading to updates of our model of the world. This has been translated to more general frameworks of computational neuroscience: Integration of prior object knowledge and sensory information is thought to occur through recurrent feedforward and feedback loops in the visual cortex, which allow us to make Bayes-optimal inferences based on the relative precision of the sensory information and prior expectations (Lee & Mumford, 2003). Ernst and Banks (2002) found that when sensory information is less precise, prior expectations have a greater influence on perception, whereas more precise sensory information leads to less influence of prior expectations. In contrast, more precise sensory information leads to less influence of prior expectation on perception. This leads to the notion of the brain as a statistical machine, as has been proposed by Dayan et al. (1995) who introduced the brain as a *Helmholtz Machine*. Following early

ideas by Helmholtz (1867), Dayan et al. (1995) describe the human perceptual system as a "statistical inference engine whose function is to infer the probable causes of sensory input" (Dayan et al. 1995, p.889).

The *Bayesian Brain* postulates the integration of prior knowledge and sensory input in the brain as a crucial aspect of our ability to make accurate predictions about the world around us. The brain is always testing its models of the world – like recurrent hypothesis tests, expectations, knowledge, and assumptions are tested against sensory information, which either confirm, alter or disregard our model of the world – and ultimately lead to learning, i.e. a refinement of our model of the world. By understanding how this integration occurs, we can gain insights how the brain processes information and makes inferences.

In the next section, we will explore how the integration of prior knowledge and sensory input can be incorporated into a Bayesian framework for understanding the perception of pain in a *Bayesian Pain Model.*

### 3.1 The Bayesian Brain in Pain: The Bayesian Pain Model

The idea that the brain processes sensory information in a Bayes-optimal manner has been used to explain the phenomenon of placebo hypoalgesia, in which positive expectations about pain can lead to a decrease in pain perception, and nocebo hyperalgesia, in which negative expectations can lead to an increase in pain perception (Anchisi & Zanon, 2015; Büchel et al., 2014; Ongaro & Kaptchuk, 2019; Wiech, 2016). In a *Bayesian Pain Model*, placebo hypoalgesia and nocebo hyperalgesia are explained by a Bayes-optimal integration of top-down prior expectations (i.e., prediction of pain and pain relief) with bottom-up sensory signals at different points in the neural hierarchy (Büchel et al., 2014).

An analytical Bayes-optimal solution to the integration of top-down prior pain expectations and bottom-up sensory input can be achieved by using (biologically plausible) approximate Gaussian distributions of these parameters (see Laplace approximation; Friston et al., 2007; Friston & Penny, 2011). Gaussian distributions of the prior, representing pain expectations, and the likelihood, representing sensory information, allow for an analytical integration to estimate the Gaussian posterior, representing the pain percept (see Eq. 2). Normal distributions can be described by two parameters, the mean $\mu$ of the distribution and the precision of the distribution $\sigma^2$. In this view, the posterior is dependent on the mean of the intensity expectation ($\mu_{prior}$)

and its precision (standard deviation; denoted as $\sigma^2_{prior}$) as well as the mean intensity of the sensory information (nociceptive signal; $\mu_{likelihood}$) and the precision of the sensory information ($\sigma^2_{likelihood}$), which is formalized in the following equations, representing the mean and the precision of the resulting posterior distribution:

(Eq.2)

$$\mu_{posterior} = \frac{\mu_{prior} * \sigma^2_{\ likelihood} + \mu_{likelihood} * \sigma^2_{\ prior}}{\sigma^2_{\ likelihood} + \sigma^2_{\ prior}}$$

$$\sigma^2_{posterior} = \frac{\sigma^2_{likelihood} * \sigma^2_{prior}}{\sigma^2_{likelihood} + \sigma^2_{prior}}$$

The following examples show how the integration of expectations and sensory information can explain phenomena like placebo hypoalgesia and nocebo hyperalgesia. These examples are adapted from the work of Büchel et al. (2014).

In the first example, imagine that a noxious stimulus is associated with a pain intensity of 50 VAS on a 0-100 point Visual Analogue Scale (VAS). This is considered the control case, meaning that there are no expectations involved: In Bayesian integration, this is represented by a flat prior. The intensity of the pain stimulus is encoded in the mean of a Gaussian distribution, while the sensory precision is represented by the standard deviation, which indicates the width of the distribution. In this case, the flat prior does not change the mean when it is combined with the likelihood, so the resulting perception (posterior) is the same as the sensory input in the mean (Figure 1a).

In the second example, imagine that a particular cue is repeatedly associated with a pain stimulus of 30 VAS, leading to conditioning (Figure 1b). Instead of the 30 VAS stimulus, a 50 VAS stimulus is then presented, as in the control case described above. In this case, the prior includes the conditioned expectation of 30 VAS, which is lower than the sensory information of 50 VAS. This represents a placebo expectation, because the expected pain is less than the actual pain stimulus. In Bayesian integration, the combination of the prior and the likelihood produces a posterior (perception) that falls between the two. In this example, this leads to placebo hypoalgesia (compared to the control case), because the pain is now perceived as 40 VAS rather than 50 VAS.

In contrast, the third example considers the case of nocebo hyperalgesia (Figure 1c). In this case, a particular cue is repeatedly associated with a pain stimulus of 70 VAS. Instead of the 70 VAS stimulus, a 50 VAS stimulus is presented, as in the control case. This time, the prior corresponds to the conditioned expectation of 70 VAS, which is higher than the sensory information of 50 VAS. This represents nocebo expectations, because the expected pain is greater than the actual pain stimulus. In Bayesian integration, this leads to nocebo hyperalgesia, where the pain is perceived as higher than it actually is, as in this example, where it is perceived as 60 VAS.

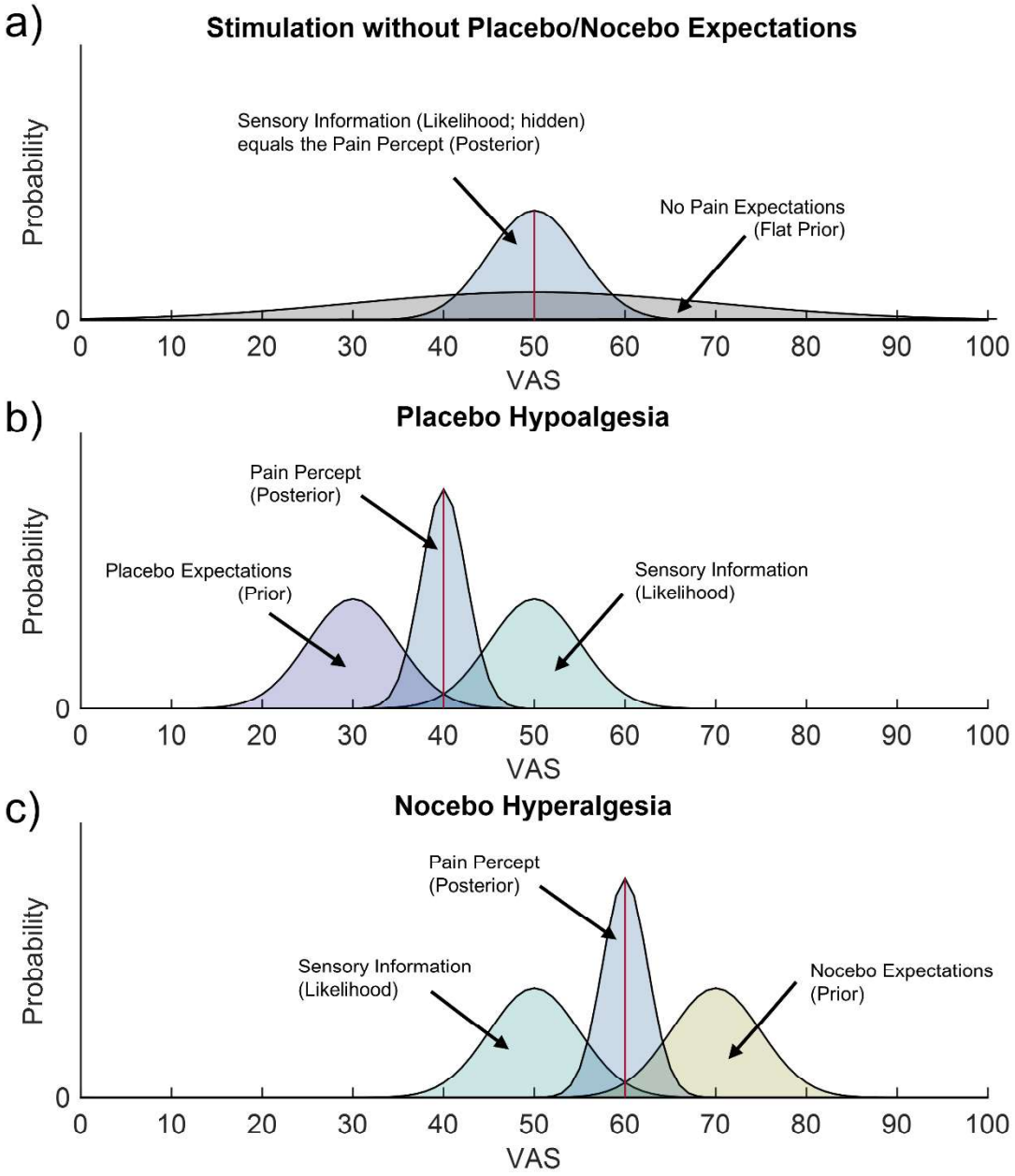



**Figure 1.** The *Bayesian Pain Model* in placebo hypoalgesia and nocebo hyperalgesia. Adapted from Büchel et al. (2014). Gaussian distributions characterize nociceptive sensory input (likelihood; green), placebo expectations (prior; purple), nocebo expectations (prior; yellow) and the pain percept (posterior; blue). The red line indicates the mean of the posterior distribution. In the first example (a), no expectations were generated, represented by a (flat) prior centered on the mean of the stimulus. The pain percept (posterior; blue) is identical to the nociceptive stimulus (likelihood; hidden). The *Bayesian Pain Model* explains the modulations of the pain percept (posterior; blue) in (b) placebo hypoalgesia and (c) nocebo hyperalgesia by the Bayes-optimal integration of prior experiences (here centered at VAS = 30 for placebo expectations and at VAS = 70 for nocebo expectations) with incoming nociceptive information (likelihood; green). In the second example, representing placebo hypoalgesia (b), placebo expectations (prior; purple) were generated, which in Bayesian integration with the nociceptive stimulus (likelihood; green) leads to a placebo effect and a pain percept (posterior; blue) shifted towards lower VAS ratings. In the third example, representing nocebo hyperalgesia (c), nocebo expectations (prior; yellow) were generated, which in Bayesian integration with the nociceptive stimulus (likelihood; green) leads to a nocebo effect and a pain percept (posterior; blue) shifted towards higher VAS ratings.

Bayesian integration thus requires the intensity of expectations and the intensity of sensory information, which are weighted by their precision – together they constitute a

pain percept. The influence of precision has been demonstrated empirically by manipulation of the level of precision of prior treatment expectations: expectation-based effects are more pronounced with more precise treatment expectations (Grahl et al., 2018).

In the following chapters, I will use the term *Bayesian Pain Model* to refer to the Bayesian integration of prior information and sensory input to form a pain percept, as delineated in this chapter with the example of placebo hypoalgesia and nocebo hyperalgesia. This *Bayesian Pain Model* describes the optimal integration of expectations and sensory information based on the simplification of information using Gaussian distributions. This allows for the analytical solutions of Bayesian integration. In order to provide a comprehensive framework for understanding these processes in the brain, I will briefly discuss the *Free Energy Principle* and *Predictive Coding* as theoretical frameworks for further hypotheses developed in my research in the next section.

## 3.2 The Free Energy Principle

The models of neural processing discussed in this chapter are based on the idea that internal models are constantly compared to sensory input and updated based on the processing of prediction errors during learning. Friston et al. (2006) have outlined this concept in the *Free Energy Principle*, which can be used to explain a wide range of cognitive, affective, and evolutionary processes (Brown et al., 2013, 2013; Friston, 2010; Friston et al., 2006; Friston & Kiebel, 2009; Kaplan & Friston, 2018; Parr et al., 2022). The *Free Energy Principle* is rooted in the concept of *Predictive Coding,* and posits that any self-organizing system in equilibrium with its environment will strive to minimize its free energy (Friston, 2010). By reducing prediction errors, systems can maintain their order and develop models that make more accurate predictions about the sensory world, leading to more efficient information encoding (Friston, 2010; Friston & Kiebel, 2009).

In other words, the brain is constantly seeking to create an accurate representation of the world. To do this, it generates models of the world and updates them when divergent information is received. However, Friston et al. (2006) describe another way to reduce prediction errors and improve the match between internal models and sensory input through the use of *Active Inference* (Brown et al., 2013; Friston et al., 2006, 2009, 2013, 2017). The brain continuously tries to reduce the prediction errors

of its generative model of the world, and in addition to updating internal models through learning, it can also minimize prediction errors through action. Thus, the minimization of prediction errors can be achieved in two ways: by refining predictions (i.e. updating our model of the world) to match sensory input, or by performing an action that brings the world into alignment with our predictions.

In the context of *Active Inference*, it has been suggested that the precision of sensory information about the consequences of one's actions must be attenuated in order for these actions to occur (Brown et al., 2013). This is because, according to the *Free Energy Principle*, action is a way of fulfilling expectations: for example, one might expect their arm to be in a certain location, and as a result, they move their arm to that location to reduce the mismatch between their expectation and the actual location of their arm at the original position. Without reducing the precision of sensory information during action, this expectation of the new arm location would be overruled by prediction errors arising from the sensory information about the arm's actual location.

In essence, the *Free Energy Principle* is built on a *Predictive Coding* view of the brain, as both expectations and prediction errors are crucial for minimizing free energy and updating the generative model of the world (Friston & Kiebel, 2009). *Active Inference* is an extension of the *Free Energy Principle*, which suggests that the brain can also minimize prediction errors through action.

In the following chapter, we will explore the concept of *Predictive Coding* in greater detail, as it forms the foundation of our hypothesis about pain processing in the context of the *Bayesian Pain Model*.

## 4.0 Predictive Coding

Early accounts of *Predictive Coding* were developed within the framework of computational models of inhibition processes in the retina and visual cortex (Rao & Ballard, 1999; Srinivasan et al., 1982). For example, it was proposed that the antagonistic surround of a receptor could be used to calculate the weighted mean of signals from neighboring receptors, allowing the generation of a statistical prediction of the signal at the center (Srinivasan et al., 1982). This prediction is based on the intensity values of the surrounding regions and is combined with the sensory information entering the particular point. This process can be understood in Bayesian

terms as the combination of a prior (the statistical estimate of the intensity at the point) with a likelihood (the sensory information) to generate a transmission signal.

*Predictive Coding* principles also suggest that the hierarchical organization of the cortex plays a role in how higher brain areas influence the processing of lower brain areas. There is evidence of a hierarchical organization in the visual system, in which primary regions like the primary visual cortex receive sensory input, while higher areas are involved in associative functions (Maunsell & van Essen, 1983; Zeki & Shipp, 1988). The brain is thought to use *Predictive Coding* processes to model complex scenarios involving internal representations combined with sensory information. Most of the experimental data supporting the *Predictive Coding* framework have come from studies of the visual system (Egner et al., 2010; Hesselmann et al., 2010; Jehee & Ballard, 2009; Sterzer et al., 2008; Strube et al., 2021b; Summerfield et al., 2006; Uran et al., 2022).

*Predictive Coding* mechanisms have also been used to explain processes in other sensory modalities, such as audition (Arnal & Giraud, 2012; Lesicko et al., 2022; Moran et al., 2013; Todorovic et al., 2011; Xuejing & Xin, 2019; but see Heilbron & Chait, 2018 for opposing views), olfaction (Zelano et al., 2011; Zhao et al., 2021), and interoception (Barrett, 2017; Barrett & Simmons, 2015; Seth et al., 2012), as well as higher-order cognitive processes such as the perception of causality (van Pelt et al., 2016).

In the next step, we will delve into the application of *Predictive Coding* to the study of pain perception. We will explore how *Predictive Coding* principles can be used to understand the mechanisms underlying pain processing in the brain.

### 4.1 Predictive Coding in Pain

The ascending and descending pathways in the brain involved in pain processing resemble a recurrent system that allows for *Predictive Coding* (Büchel et al., 2014). In this system, bottom-up sensory information is transmitted via the ascending pathway, while expectations are encoded in the descending pathway. This allows the brain to continually compare its expectations of pain with incoming sensory information, updating these expectations as needed to minimize prediction errors and optimize the expectations of future pain.

*Predictive Coding* mechanisms have been investigated in the context of pain, and have been linked to activity in the anterior insula, which exhibits the expected response

pattern of *Predictive Coding* with encoded expectations and prediction errors. The posterior insula and somatosensory cortex, on the other hand, are associated with the representation of stimulus intensity (Fazeli & Büchel, 2018; Geuter et al., 2017). The posterior insula has also been shown to be involved in the correlation of brain activation with higher-than-expected intensity, known as signed prediction errors (Horing & Büchel, 2022). These findings provide strong evidence that *Predictive Coding* plays a key role in pain perception

The posterior insula receives direct input from the spinothalamic tract, which is involved in transmitting sensory information in somatosensation and pain perception (Craig, 2002; Dum et al., 2009). This region is also functionally connected to the somatosensory cortices (Wiech et al., 2014). In the context of *Predictive Coding*, one could theorize that the posterior insula integrates sensory information with cognitive factors such as expectation to compute prediction errors in the anterior insula (Horing & Büchel, 2022). Importantly, these response patterns in the anterior insula have been shown to be independent of the aversiveness of the stimuli (Fazeli & Büchel, 2018), and to be modality-unspecific: the anterior insula correlated with absolute prediction errors in both aversive auditory and painful stimulation, indicating a general aversive surprise signal (Horing & Büchel, 2022). This suggests that the *Predictive Coding* mechanisms underlying pain perception are not specific to the sensory modality involved, but instead reflect a more general process of surprise minimization.

The *Predictive Coding* framework proposes that the descending and ascending pathways involved in pain processing integrate sensory information with cognitive factors such as expectation to compute prediction errors. The anterior insula is a key region for this process, as it has been shown to exhibit response patterns that are consistent with the predictions of *Predictive Coding*, particularly with respect to absolute prediction errors.

In the next section, we will explore the role of oscillatory patterns in *Predictive Coding*, with a focus on canonical microcircuits (Bastos et al., 2012).

### 4.2 Microcircuits in Predictive Coding

*Predictive Coding* is theorized to be associated with canonical cortical microcircuits where neuronal populations are associated with specific computational roles (Bastos et al., 2012). It is postulated that feedforward prediction errors are projected from lower cortical levels to higher granular levels (Bastos et al., 2012). From there, they are

transmitted to excitatory and inhibitory interneurons in supragranular layers, where expectations are thought to be encoded. Extrinsic feedback connections originate largely from superficial pyramidal cells while feedforward connections largely originate from deep pyramidal cells (Felleman & Van Essen, 1991).

These layers of the cortex are associated with asymmetry in the properties of their oscillatory power spectra: Supragranular sites have higher broadband gamma power, while granular and infragranular layers have greater power in the alpha and beta range (Maier et al., 2010). Buffalo et al. (2011) found that the spiking activity of neurons in the superficial layers of the visual cortex is more coherent with gamma-frequency oscillations in the local field potential, while neurons in the deep layers are more coherent with alpha-frequency oscillations. These findings suggest that the different layers of the cortex exhibit distinct oscillatory patterns, which may play a role in the coordination of information flow and the computation of prediction errors in the *Predictive Coding* framework.

The distinct oscillatory patterns observed in the different layers of the cortex can be used to make specific predictions about the role of feedback and feedforward connections in *Predictive Coding*. Because feedback connections predominantly originate from deep layers, while feedforward connections originate from superficial layers, we would expect that expectations, which are thought to be mediated by feedback connections, would be expressed at lower frequencies than feedforward prediction error signals.

Arnal and Giraud (2012) theorized accordingly that *Predictive Coding* should be "characterized by alternation of gamma-forward dominant and beta-backward dominant phases" (p. 394).

In the next chapter, we will integrate our understanding of oscillatory patterns in *Predictive Coding* with the existing literature on oscillatory activity in pain processing. We will briefly review the evidence for oscillatory patterns in pain perception, as well as the role of top-down modulation in shaping these patterns.

## 5.0 Oscillatory Patterns of Pain Processing

As discussed in the previous chapter, *Predictive Coding* is associated with specific oscillatory patterns (Arnal & Giraud, 2012; Bastos et al., 2012). However, many studies investigating pain perception have relied on fMRI, which has the disadvantage of a low

temporal resolution of brain signals (Ogawa et al., 1992). This means that it is difficult to make precise statements about oscillatory activity or the exact temporal sequence of signals using fMRI data. To overcome this limitation, it may be necessary to use other techniques such as electroencephalography (EEG) or magnetoencephalography (MEG) to measure oscillatory activity more accurately, and to characterize its relationship to pain perception.

EEG and MEG studies have consistently shown that oscillatory activity in the brain varies in response to different intensities of painful stimuli, with changes observed in a range of frequency bands, including infraslow (below 0.1Hz), theta (3-8Hz), alpha (8-12Hz), beta (12-30Hz), and gamma (>30Hz) oscillations (see Ploner et al., 2017 for a comprehensive review).

In a study by Schulz et al. (2015), researchers measured the brain's response to two different types of pain: tonic and phasic. They found that when the brain was subjected to subjective changes in tonic pain, which in this case was a continuous, 10-minute-long-lasting pain, there was an increase in gamma activity in the medial prefrontal cortex in close proximation to premotor and cingulate cortices (Schulz et al., 2015), while typically, brief painful stimuli have been shown to induce gamma oscillations in somatosensory cortices. Moreover, in tonic pain, stimulus intensity is inversely related to beta and alpha oscillations (Chen & Rappelsberger, 1994; Ferracuti et al., 1994; Peng et al., 2014; Schulz et al., 2015).

The specific pain protocol used in a study can have a significant impact on the representation of pain in the brain. This is because besides its amplitude, the temporal properties of the pain stimulus can change the way it is encoded and processed in the brain. For example, the representation of pain in the brain may be different for a brief, phasic pain stimulus compared to a longer-lasting, tonic pain stimulus. It is therefore important to consider the specific pain protocol used in a study when interpreting the results, as this can affect the way pain is represented in the brain (Horing et al., 2019).

In the next section, we will examine the evidence for top-down modulation of oscillatory patterns in pain perception. Previous studies have shown that oscillatory activity in pain perception is not solely determined by the intensity of the nociceptive stimulus, but can also be influenced by factors such as expectations, attention, and context (Ploner et al., 2017). This suggests that there may be a role for top-down modulation of pain-

related oscillatory activity, in which higher-level cognitive processes influence the way pain is encoded and processed in the brain.

### 5.1 Oscillatory Patterns of Top-Down Pain Modulation

The *Flexible Routing model* of pain perception, proposed by Ploner et al. (2017), suggests that oscillatory patterns of pain can be influenced by top-down factors such as expectations or attention. This is thought to be linked to alpha/beta oscillations and synchronization of alpha/beta activity across different brain areas, whereas gamma oscillations are proposed to be involved in feedforward signaling. Previous research has explored the spectral properties of mechanisms involved in pain perception, and has found evidence for a relationship between alpha/beta oscillations and top-down modulation of pain. For example, one study using MEG found that alpha suppression in the anterior insula is linked to the expectation of pain in a situation where painful and non-painful stimuli were interspersed (Franciotti et al., 2009). Another study suggested that alpha desynchronization in response to predictable painful stimuli may be a neural marker of attentional preparation (Babiloni et al., 2003).

Ohara et al. (2006) conducted an experiment in which they used subdural electrodes to measure the neural activity of two subjects in the primary somatosensory, perisylvian, and medial frontal cortex. The subjects were exposed to laser pulses that induced acute phasic pain, and their brain activity was monitored during two different conditions. In one condition, the subjects were asked to pay attention to the laser stimulus by performing a cognitive task, while in the other condition, they were distracted by a magazine article. The researchers found that attention increased the functional connectivity of brain activity in the beta range between the primary somatosensory cortex and the perisylvian cortex before the laser stimulus was applied. Additionally, they observed an increase in synchronization in the alpha range in the primary somatosensory cortex and the medial frontal cortex (Ohara et al., 2006).

In a similar experiment by Liu et al. (2011) with three subjects, causal influences were measured between the primary somatosensory cortex, the perisylvian cortex, and the medial wall during an attention-distraction task. Oscillatory activity in the alpha and beta ranges were found to be important for the attentional modulation of pain (Liu et al., 2011). In a study by Hauck et al. (2015), results showed that attention towards pain led to a decrease in alpha and an increase in gamma activity in the insula. In a similar study, attentional augmentation of pain processing was suggested based on increased

gamma band activity by directed attention leading to enhanced saliency (Hauck et al., 2007). May et al. (2012) report that spatial attention towards pain resulted in a modulation of pre- and post-stimulus alpha activity, which in summary supports the idea of alpha activity in top-down control mechanisms of pain (May et al., 2012).

### 5.2 Oscillatory Patterns of Pain Expectation

According to the *Predictive Coding* framework and the *Free Energy Principle*, expectations play a key role in pain processing. In a study by Albu and Meagher (2016), when participants expected pain to be worse, they actually experienced more pain and showed increased activity in the low alpha frequency range. This increase in alpha activity was associated with negative thoughts and emotions about the pain, such as pain catastrophizing, rather than the pain intensity itself, suggesting a cognitive-affective modulation via alpha synchronization (Albu & Meagher, 2016). In another nocebo study by Thomaidou et al. (2021) beta-band activity in the brain was associated with the magnitude of the nocebo response. Individuals with strong long-range temporal correlations in the beta band had larger nocebo responses than those with weaker activity (Thomaidou et al., 2021). The researchers also found that alpha power was reduced during the acquisition of the nocebo response, and that alpha power was higher in nocebo-augmented pain compared to the baseline (Thomaidou et al., 2021). In a study involving placebo treatment, alpha activity increased significantly post-placebo-treatment (Huneke et al., 2013).

In a study by Nickel et al. (2022), participants were presented with brief painful stimuli of low and high intensities that were probabilistically cued. The researchers found that the expectations of the participants regarding the upcoming pain influenced their brain activity in the alpha and beta frequency ranges before the presentation of the pain stimulus. However, during the painful stimulation, the researchers did not observe any modulation of oscillatory power due to expectations (Nickel et al., 2022). Babiloni et al. (2003) found that anticipatory alpha activity was associated with the subjective evaluation of pain intensity. Additionally, it has been observed that distracted participants during an anticipatory period reported lower pain intensity and unpleasantness, accompanied by a reduction in alpha activity in the fronto-central midline region of the brain (Del Percio et al., 2006). These findings suggest a potential role for oscillatory activity in the alpha and beta frequency ranges in the modulation of pain perception by expectations. This effect is thought to be related to top-down modulation of pain processing, and is consistent with theories of *Predictive Coding*,

which suggest that the brain uses expectations to predict sensory input and generate prediction errors.

In summary, the theoretical foundation of my studies is based on understanding the brain as a statistical machine that uses Bayesian principles to integrate prior expectations and sensory information. The *Bayesian Br*ain can be explained using theories like the *Free Energy Principle* and its extension through *Active Inference* and *Predictive Coding*, which can be applied to the recurrent system of pain processing in the brain with its ascending bottom-up pathways and descending top-down pathways. These processes can be measured through oscillatory patterns using EEG, allowing for the investigation of pain and its top-down processing.

## 6.0 Testing Predictive Coding, Active Inference, and the Bayesian Pain Model

In this chapter, I will present the empirical studies that were conducted as part of this cumulative dissertation. These studies were designed to investigate the role of *Predictive Coding* and top-down modulations in the neural processing of pain, with a focus on oscillatory patterns. I will describe the methods and results of these studies and discuss their implications in the light of a *Bayesian Pain Model*.

The first study (Reprint 15.1; Strube et al. 2021a) is a replication study using EEG to investigate the implementation of *Predictive Coding* in the cortical pain network. Here, we tested the specific hypothesis that expectations should be encoded in lower frequencies than prediction errors. The high temporal resolution of EEG allowed for the separation of expectation and prediction error signals, which was not possible in previous fMRI studies. The study found that alpha-to-beta synchronization occurred immediately after a predictive cue, which was associated with higher intensity expectations. In contrast, prediction errors were encoded in the gamma range during phasic painful thermal stimulation.

The second study (Reprint 15.2; Strube et al. 2021b) examines the use of *Predictive Coding* for negative affective image stimuli. The findings indicate that there are different patterns of neural activity for this type of stimuli, suggesting the existence of modality-specific oscillatory networks. In contrast to our observations for painful stimuli, we found that the affective valence, valence expectations, and prediction errors were encoded in low frequency bands during the presentation of aversive pictures. This

suggests that the signals are specific to the modality being processed, and that the orchestration of *Predictive Coding* differs between pain and affective visual processing.

The third study (Reprint 15.3; Strube et al. (2022), submitted) explores how the contextual aspect of agency can be integrated into a *Bayesian Pain Model* and whether this supports the predictions of the *Active Inference* model. We conducted a placebo/nocebo experiment in which treatment was self- or externally-initiated. The results showed that agency can modulate the effectivity of treatment and that this can be explained as a shift in intensity expectations.

## 6.1 Study 1: Predictive Coding in Pain

Strube, A., Rose, M., Fazeli, S., & Büchel, C. (2021). The Temporal and Spectral Characteristics of Expectations and Prediction Errors in Pain and Thermoception. eLife, 10, e62809.

### 6.1.1 Introduction

In the first study, our aim was to investigate the oscillatory properties of *Predictive Coding* signals in pain processing. We specifically hypothesized that an expectation signal would be encoded in oscillatory patterns earlier than a prediction error signal. Additionally, we predicted that expectations would be represented in lower frequencies than prediction errors.

Noxious thermal stimulation of varying intensity was unreliably cued to generate expectations regarding the stimulation. I.e., a predictive cue was associated with low (42°C), medium (46°C) or high pain (48°C), where a mismatch of expectation and stimulation has been conceptualized as a prediction error.

Geuter et al. (2017) and Fazeli & Büchel (2018) conducted similar studies on the *Predictive Coding* framework in relation to pain using fMRI. Geuter et al. (2017) used a cued pain design with two pain stimulus intensities and observed anterior insula responses that followed the response patterns predicted by the *Predictive Coding* framework, indicating the involvement of expectations and prediction errors. Fazeli & Büchel (2018) extended this work by using three pain intensities and a second modality to control for aversiveness. They found that the ventral anterior insula represented pain intensity, expectation, and absolute prediction errors, and that this could not be explained by aversiveness. Both studies showed that the insula plays an important role in the *Predictive Coding* model, but were unable to establish its temporal aspects. This

motivated the current study, which uses EEG to further investigate the temporal and spectral aspects of these signals.

In this study, we replicated the design of Fazeli & Büchel (2018) with slight modifications for EEG measurements and tested the hypothesized spectral patterns of the microcircuits involved in *Predictive Coding* using time-frequency analysis of EEG data. As discussed in chapter 4.2, it has been suggested that prediction errors should be expressed by higher frequencies than the predictions that generate them (Arnal & Giraud, 2012; Bastos et al., 2012).

### 6.1.2 Methods

The current EEG study was designed to experimentally manipulate expectations and prediction errors using preceding cues containing unreliable information about the subsequent pain stimulus. This was achieved by the presentation of predictive cues (presented for 1000-1400ms) preceding a painful heat stimulus (42°C, 46°C and 48°C für low, medium and high stimulus intensities, respectively). Thermal stimulation was applied using a 30 × 30mm$^2$ Peltier thermode (CHEPS Pathway, Medoc, Israel). As a control condition (see 6.2), we included aversive emotional picture stimuli from the International Affective Picture System (IAPS; Bradley & Lang, 2007). The color of the cue (triangle) indicated unreliably the modality of the stimulus (orange for picture and blue for heat). A white digit written inside of each triangle indicated unreliably the intensity of the subsequent stimulus (a 1, 2 and 3 for low, medium and high intensity). Based on this cue-stimulus configuration, we could reliably produce expectations and prediction errors with control of contingencies between cues and stimuli. The intensity was cued correctly in 60% of all trials, whereas the modality was cued correctly in 70% of all trials.

This cued pain paradigm allowed us to directly examine the spectral patterns associated with prediction errors and expectations. After conducting a behavioral training session to ensure that participants understood the contingencies, we recorded their EEG while they continuously rated cued pain and picture trials. After artifact rejection, correction via independent component analysis (ICA; Jung et al., 2000; Makeig et al., 1996) and time-frequency transformations, we performed explorative cluster permutation tests to assess associations with stimulus intensity, expectations and prediction errors in cue- and pain-locked EEG time-frequency data (see Reprint 15.1., methods for details).

For the analysis of EEG data, we decided to correct for multiple comparisons using non-parametrical permutation tests of clusters of neural activity (Maris & Oostenveld, 2007). We based our hypotheses about *Predictive Coding* on the functional architecture of canonical cortical microcircuits, but previous studies did not provide enough detailed information about the timing and frequency patterns of these processes. We used cluster permutation tests to account for the fact that biological processes do not always occur at a specific frequency or time point, and that multiple electrodes may detect activity at the same time.

### 6.1.3 Results

In this study, we used cluster permutation tests to analyze the time-frequency patterns of EEG data during a cued pain paradigm with three different stimulus intensities. Our findings show that there were clear differences in activity between the different levels of aversiveness. We found multiple clusters of activity in the theta (3-8Hz), alpha-to-beta (8-30Hz), and gamma (>30Hz) bands that were related to the strength of the pain stimulus. Behavioral data also showed that the cued intensity influenced pain perception, i.e. higher pain expectations were associated with higher pain ratings.

Importantly, our results indicated that temporally and spectrally separable clusters of oscillatory activity are associated with components of *Predictive Coding*. One early, low-frequency cluster (3-30Hz) was specifically linked to anticipation in pain perception, or cued intensity. A later-appearing cluster at higher frequencies (31-100Hz) was linked to negative absolute prediction errors in pain. On a behavioral level, pain ratings were associated with expectations; higher pain expectations were associated with higher pain ratings.

### 6.1.4 Discussion

In this study, we found that expectation and prediction errors have a significant impact on pain perception and oscillatory processes in the brain. The modulation of behavioral pain ratings by expectation (i.e. higher expectations are associated with higher pain) is supporting the *Bayesian Pain Model*, where higher expectations would be modelled via a higher prior mean. In Bayesian integration this would result in higher pain ratings with higher expectations (Büchel et al., 2014).

We observed a negative modulation of gamma activity by absolute prediction errors, i.e. lower gamma activity was measured when the predictive cue did not match the intensity of the upcoming stimulus during pain. This is contrast to prediction error

effects in the visual (Bauer et al., 2014; van Pelt et al., 2016) and auditory (Edwards et al., 2005; Parras et al., 2017) domains. Interestingly, while not typically discussed in the framework of *Predictive Coding*, other cognitive domains show this pattern of decreased gamma activity with mismatch. For example, it has been shown that that gamma power is associated with successful matching (representing the absence of a prediction error) of external input with internal representations (Herrmann et al., 2004; Osipova et al., 2006; Wang et al., 2018), and in this context, gamma band activity have been discussed in terms of representing a match between bottom-up and top-down information (Herrmann et al., 2004). One further example is increased gamma activity with higher cloze probabilities in language comprehension (Hald et al., 2006; Molinaro et al., 2013; Obleser & Kotz, 2011; Wang et al., 2018), where a critical word that is semantically predictable by the preceding sentence (and is thus related with a high cloze probability) leads to higher gamma activity as compared to words which are semantically unpredicted (low cloze probability), which would be represented by a higher prediction error in a *Predictive Coding* model.

In this study, we were able to identify the temporal orchestration of essential parameters of the *Predictive Coding* process in spectral patterns of neural activity. Firstly, we found that expectations were associated with increased alpha-to-beta activity following a predictive cue. In a *Bayesian Pain Model*, this would represent prior information that would then be integrated with sensory information during painful stimulation. We observed that this prior information modulated pain perception accordingly, with higher expectations being associated with higher pain ratings, as predicted by the *Bayesian Pain Model* (Büchel et al., 2014). Secondly, we found that a decrease in gamma activity was associated with prediction errors during painful stimulation – this is atypical and was not predicted by theoretical accounts of the architecture of microcircuits, which posit an increase in gamma activity in response to prediction errors. These findings provide new insights into the temporal dynamics of *Predictive Coding* in the perception of pain.

### 6.2. Study 2: Predictive Coding in (Aversive) Visual Perception

Strube, A., Rose, M., Fazeli, S., & Büchel, C. (2021). Alpha-to-Beta and Gamma-Band Activity Reflect Predictive Coding in Affective Visual Processing. Scientific Reports, 11(1), 1-15.

### 6.2.1 Summary

Affective image stimuli were also presented as part of the paradigm of the first experimental study. We have published the results of the analysis of these data in another publication (Reprint 15.2; Strube et al., 2021b).

I also refer to the publication (Reprint 15.2; Strube et al., 2021b) for a detailed classification of the results in the context of affective picture processing. Since the scientific work presented in the context of this dissertation is essentially about pain processing, I reduce the content of this section to the relevance of this study in the context of *Predictive Coding* in the pain domain. Ultimately, the visual-affective modality was introduced to serve as a control condition for the pain condition to delineate whether different mechanisms are responsible in *Predictive Coding* in pain compared to *Predictive Coding* in a different (aversively stimulated) modality, which I will discuss in this chapter. These results are useful to assess whether the spectral and temporal orchestration of *Predictive Coding* we found in the first study are specific to pain stimuli or whether we find comparable results for affective picture stimuli.

The affective picture stimuli were presented in the same way as the pain stimuli in the first study with the same contingencies. Since both pain stimuli and picture stimuli were presented to the same group of subjects and the same EEG preprocessing and analysis techniques were used, results are comparable and differences can be attributed to the use of the different stimulus materials (see Reprint 15.2; Strube et al. 2021b for detailed methods).

Affective visual stimuli are usually accompanied by a desynchronization of alpha activity (8-12 Hz; Cui et al., 2013; De Cesarei & Codispoti, 2011; Schneider et al., 2018; Schubring & Schupp, 2019, 2021). In the context of *Predictive Coding*, this could be interpreted as an expectation signal encoded in the increase of alpha activity. Moreover, affective imagery is associated with event-related synchronization of gamma activity (>30Hz; Boucher et al., 2015; Güntekin & Tülay, 2014; Keil et al., 2001; Martini et al., 2012; Müller et al., 1999; Schneider et al., 2018). While expectation signals are associated with low frequencies, a modulation of gamma would be postulated for the transmission of feed-forward prediction errors based on canonical microcircuits theorized for *Predictive Coding* (Arnal & Giraud, 2012; Bastos et al., 2012).

This yields specific hypotheses for *Predictive Coding* in affective visual image processing: desynchronization of alpha activity and synchronization of gamma activity typically associated with affective stimuli signal components of *Predictive Coding*. I.e., alpha activity should be modulated by expectations, whereas gamma activity should be modulated by prediction errors.

The paradigm used in this study is a probabilistic cue paradigm in which the contingency of cues and valence intensity of affective imagery are learned in a training session. Also, cue-stimulus contingencies did not change during the experiment. In accordance with the pain levels of the first study (see 6.1; Reprint 15.1; Strube et al. 2021a), three affective picture levels were presented to the participants.

The valence of these pictures was clearly discriminable, as behavioral ratings and EEG time-frequency patterns in the alpha-to-beta (8-30Hz) band were linearly associated with our valence intensity manipulation. Valence ratings were further modulated by prediction errors: a larger mismatch between expected and actual stimulus led to larger valence ratings.

Initially, we had hypothesized that low frequencies should be modulated by expectations. Here, we found a negative association of alpha-to-beta (8-30Hz) activity and expected valence during the presentation of the image stimulus. Interestingly, we found no evidence for such a signal shortly after cue presentation before the picture stimulus appeared (i.e. a cue was presented, and after a certain lag, the picture stimulus was presented). This contrasts with our expectation signal in the pain modality: the expectation of pain modulated oscillatory patterns in an anticipatory phase before the pain stimulation occurred, immediately after the presentation of an intensity cue.

As another ingredient of *Predictive Coding*, we expected a prediction error signal in the gamma frequency band (>30Hz). The results provide evidence for this hypothesis: Changes in gamma activity were associated with absolute prediction errors. Again, there are clear differences to spectral patterns associated with *Predictive Coding* in pain. In pain stimulation, a desynchronization of gamma activity is associated with absolute prediction errors, here there is an increased gamma activity, which is typically hypothesized in *Predictive Coding* theories (Arnal & Giraud, 2012; Bastos et al., 2012).

Our findings suggest that *Predictive Coding* is involved in affective picture processing, but that oscillatory patterns are fundamentally different from the patterns involved in pain perception. By using a carefully designed paradigm that allowed us to precisely control expectations and prediction errors, we were able to observe distinct patterns of neural activity in the two different modalities. These results support the idea that there are distinct, modality-specific *Predictive Coding* processes associated with oscillatory activity in affective visual processing and pain processing.

## 6.3 Study 3: Agency in Pain Treatment and the Bayesian Pain Model

Strube, A., Horing, B., Rose, M., & Büchel, C. (2022). Agency Affects Pain Inference through Intensity Shift as Opposed to Precision Modulation [Manuscript submitted for publication]. Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf

### 6.3.1 Introduction

Expectations are fundamental in clinical pain management (Gniß et al., 2020; Peerdeman et al., 2016; Petersen et al., 2014). However, meta-analyses show that agency also plays a major role in the treatment of pain. For example, self-treatment of pain (so-called PCA: Patient-Controlled Analgesia) leads to a greater reduction in pain compared to external treatment of pain in clinical settings (Ballantyne et al., 1993; McNicol et al., 2015). The beneficial effect of agency has also been shown experimentally in pain treatment, showing modulations of pain ratings and physiological recordings (Beck et al., 2017; Helmchen et al., 2006; Jensen & Karoly, 1991; Kakigi & Shibasaki, 1992; Karsh et al., 2018; Mohr et al., 2008, 2012; Müller & Netter, 2000; Pellino & Ward, 1998; Pervin, 1963; Staub et al., 1971; Thompson, 1981; Wang et al., 2011; Weisenberg et al., 1985; Wiech et al., 2006). This process is associated with sensory attenuation (Blakemore et al., 1998, 1999, 2000; Claxton, 1975; Weiskrantz et al., 1971), namely a reduction of perceived stimulus intensity by self-production of these stimuli. Even Charles Darwin observed that “from the fact that a child can hardly tickle itself, or in a much less degree than when tickled by another person, it seems that the precise point to be touched must not be known” (Darwin, 1872). Since expectations and agency are in principle important effects in (clinical) pain treatment, the interaction of these components is another factor that deserves closer examination.

This study aimed to investigate the role of agency, as a contextual top-down factor of pain, in the *Bayesian Pain Model*. The *Bayesian Pain Model* proposes that expectations can be integrated into the model as a prior in a Bayesian integration, where sensory information is combined with prior information to form a pain percept (Büchel et al., 2014). Both the pain intensity expectation and the precision of that expectation can be modeled using Bayesian integration. These expectations are integrated with sensory information in an optimal way, leading to the formation of a pain percept. The question addressed in this study is how the influence of agency can be integrated into the *Bayesian Pain Model*.

It is interesting to consider the relationship between the principles of *Active Inference* (Friston et al., 2006, 2009, 2017) in relation to the influence of expectations and agency on pain perception. *Active Inference* is associated with a need to reduce the precision of self-produced stimuli in order to be able to act (see 3.2; Brown et al., 2013). This can be integrated into the *Bayesian Pain Model* (Figure 2), as a reduction in likelihood precision (Figure 2c). I.e., a decrease in attention to sensory information leads to less precise sensory input. In the case of self-treatment with placebo expectations (i.e. treatment expectations are better than the actual treatment), this would result in a larger weight given to placebo expectations in Bayesian integration, leading to relatively larger placebo expectation effects. In the case of self-treatment with nocebo expectations (i.e. treatment expectations are worse than the actual treatment), this would result in a higher weight given to nocebo expectations, leading to relatively larger nocebo expectation effects. This would manifest as an interaction: *Active Inference* posits that self-treatment leads to enhanced expectation effects compared to external treatment.

This is contrasted by the classical *Forward Model*, where somatosensory attenuation is explained by continuously generated predictions of the sensory consequences of a motor command. Here, accurate predictions are used to attenuate the intensity of sensory consequences of self-produced movement (Blakemore et al., 2000). This can be integrated in the *Bayesian Pain Model* as a reduction of the likelihood or prior mean – i.e., a decrease of intensity of the sensed or expected nociceptive input (Figure 2b). In self-treatment, as compared to external treatment, this would lead to a decrease in pain, regardless of placebo or nocebo expectations. In contrast to *Active Inference*,

this would not be predictive of an interaction, but of additive effects of expectations and agency.

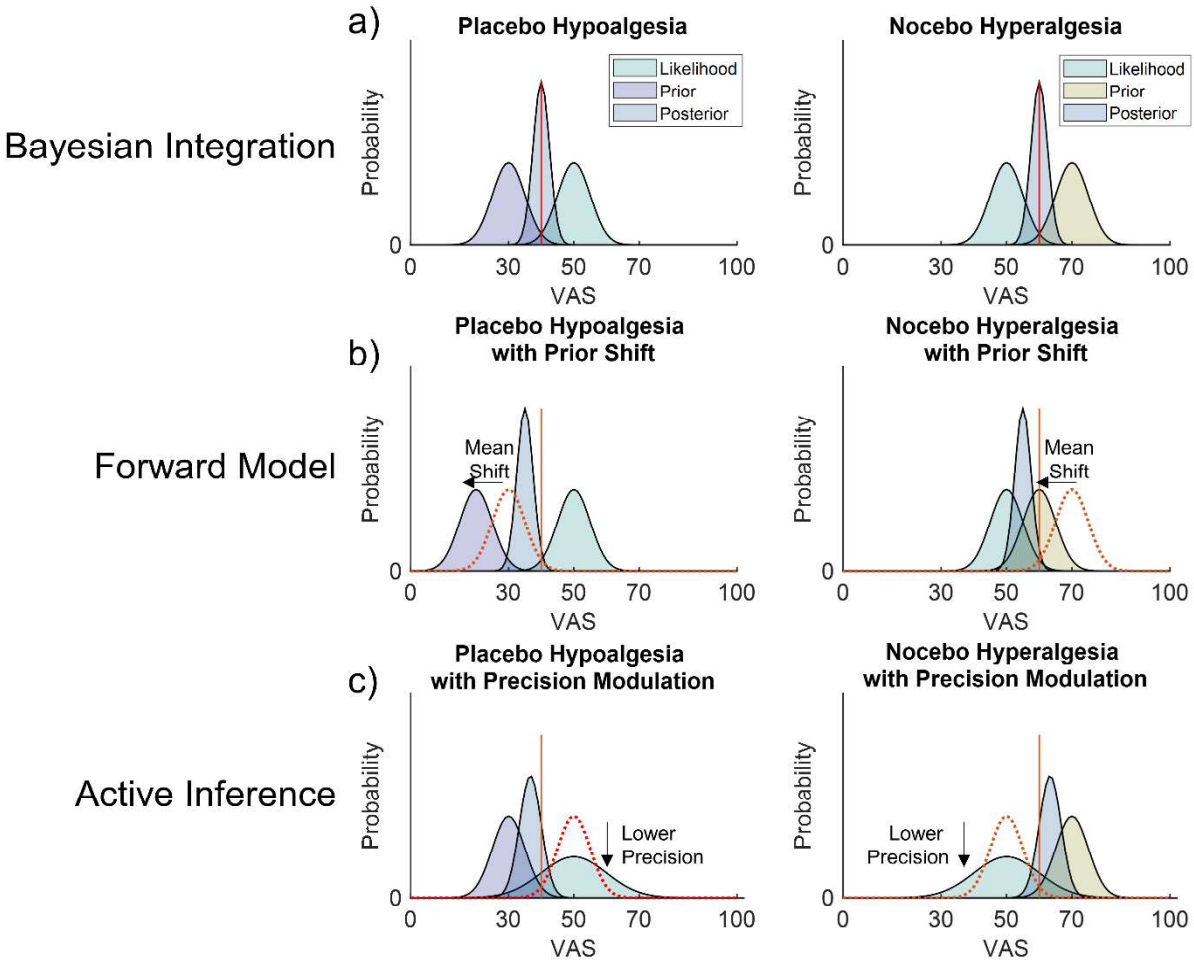



**Figure 2.** The *Bayesian Pain Model* is hypothesized to integrate agency by (b) a mean shift of prior expectations as derived from the *Forward Model* or by (c) an attenuation of sensory precision, as derived from *Active Inference.* Adapted from Strube et al. (2022, submitted). Gaussian distributions characterize nociceptive sensory input (likelihood; green), placebo expectations (prior; purple), nocebo expectations (prior; yellow) and the pain percept (posterior; blue). (a) In simple Bayesian integration in placebo/nocebo conditions without a modulation by agency, the pain percept (posterior; blue) is the result of a Bayes-optimal combination of placebo expectations (left; here centered at 30 VAS; purple) or nocebo expectations (right; here centered at 70 VAS; yellow) with nociceptive sensory information (likelihood; green). The red line indicates the mean of the posterior distribution. (b) In the *Prior Shift Model* derived from the *Forward Model,* agency over pain shifts the mean of the prior towards lower VAS values in both, placebo (left) and nocebo (right) conditions. The shift of prior distributions towards lower VAS values leads to a pain percept (posterior; blue) which is shifted towards lower VAS values as a consequence of an integration of the likelihood (green) with the shifted prior in both, placebo

hypoalgesia (left) and nocebo hyperalgesia (right). Here, the red dotted Gaussian distributions represent prior placebo/nocebo distributions without the influence of agency. The red line indicates the mean of the posterior distribution in Bayesian integration based on a prior without a shift, i.e. in conditions without agency over the pain stimulus. (c) In the *Likelihood Precision Modulation Model*, derived from *Active Inference,* agency attenuates sensory precision. The pain percept (posterior; blue) results as a Bayes-optimal combination of placebo (left) and nocebo (right) expectations with incoming precision-attenuated nociceptive information (likelihood; green). As a consequence, the relative precision of placebo/nocebo expectation increases, leading to a shift of the pain percept (posterior; blue) toward lower VAS values in placebo hypoalgesia (left) and to a shift toward higher VAS values in nocebo hyperalgesia (right), as the posterior (blue) is "drawn" towards the prior. The red dotted Gaussian distributions represent likelihood distributions without a precision attenuation by agency. The red line indicates the mean of the posterior distribution in Bayesian integration without attenuated sensory precision.

### 6.3.2 Methods

In two pain placebo experiments with self-initiated and externally initiated pain with both, placebo and nocebo conditions, we investigated how agency acted on parameters of the *Bayesian Pain Model*. We derived hypotheses of precise potential modulations of *the Bayesian Pain Model* by agency from the *Forward Model* and *Active Inference*.

In both experiments, painful stimuli were presented and participants were told that the pain would be treated subsequently by TENS (Transcutaneous Electric Nerve Stimulation). Actually, the pain was reduced by a reduction of the stimulation temperature. A TENS cover story has been shown to effectively create placebo effects and has been used in several studies (Grahl et al., 2018; Schenk et al., 2017; Thomaidou et al., 2021; van de Sand et al., 2018). Here, the TENS cover story was utilized to establish a treatment situation. In reality, no TENS stimulation was applied during the treatment of the painful stimuli. This treatment could be self-initiated by the participant or externally initiated by the experimenter (as explained to the participant; actual external treatment was initiated automatically by the computer).

This treatment was either related to high or low treatment effectivity. In conditioning trials, a cue signaling high treatment effectivity was actually associated with a stronger decrease in pain (a reduction from 70 VAS to 10 VAS) as compared to a cue signaling low treatment effectivity (a reduction from 70 VAS to 50 VAS). In test trials, both the high treatment effectivity cue and the low treatment effectivity cue were leading to the same treatment outcome (a reduction from 70 VAS to 30 VAS). Based on conditioned placebo/nocebo pain paradigms and the *Bayesian Pain Model*, we expected a

modulation of pain ratings in test trials based on placebo/nocebo expectations (Fazeli & Büchel, 2018; Geuter et al., 2013; Geuter & Buchel, 2013; Hird et al., 2018; Lorenz et al., 2005; Nickel et al., 2022; Strube et al., 2021a), i.e. expectations of highly effective treatment (placebo) should be associated with better treatment outcomes in test trials than expectations of weakly effective treatment (nocebo).

Two models have been proposed to explain the effects of agency on sensory experiences. The first main model, derived from the *Forward Model*, suggests that smaller prediction errors during self-generated movement can lead to a weaker sensation of action outcomes. Applied to the context of pain treatment, this would mean that self-treatment would result in a lower perception of pain. This idea motivated the *Likelihood Shift Model* or *Prior Shift Model* (see Figure 2b). The second main model, based on *Active Inference*, proposes that sensory attenuation is necessary for enabling action by reducing the precision of sensory evidence related to the consequences of one's own actions. This would translate to a reduction in the precision of the likelihood in the *Bayesian Pain Model*, which motivated the *Likelihood Precision Modulation Model* (see Figure 2c).

We hypothesized that if the likelihood mean is shifted by agency (as derived from the *Forward Model*), self-treatment would result in overall lower pain ratings, regardless of prior expectations. This means that the effects of agency and expectations would be additive. In contrast, if the precision of the likelihood is reduced (as derived from *Active Inference*), the impact of expectations on pain ratings would be enhanced, leading to a stronger influence of expectations on self-treatment conditions. This would manifest as an interaction between agency and expectations.

As in previous studies, we utilized cluster permutation tests to assess modulations of oscillatory activity by stimulus intensity, placebo/nocebo expectations, agency and the interaction of placebo/nocebo expectations with agency. Moreover, we tested EEG time-frequency data for correlations with between-subject placebo benefits (i.e. the difference between high and low treatment expectations in test trials) and for correlations with between-subject sensory attenuation effects (i.e. the benefit of self-treatment as derived from model parameter estimates).

The first experiment (N = 25) used continuous pain ratings to establish a precise readout of pain perception during painful heat stimulation and after treatment, whereas the second experiment (N = 54) also used EEG to assess neurophysiological

correlates of top-down modulation via expectations and agency. In both experiments, thermal pain was administered to capsaicin-sensitized skin on the left radial forearm following individual calibration to establish pain levels that were equivalent for each participant. To minimize contamination of EEG data by movement artifacts caused by continuous pain ratings through button clicks, we changed the paradigm for experiment 2 to incorporate single outcome ratings rather than continuous pain ratings. See Reprint 15.3. for detailed methods.

### 6.3.3 Results

In this study, we used three different VAS levels at treatment outcome. These levels were 10 VAS and 50 VAS for high and low treatment effectivity during conditioning trials, and 30 VAS for test trials. To determine if participants perceived these levels as different, we conducted a repeated measures analysis of variance on the final continuous rating data points (post-treatment VAS rating) from experiment 1 and on the outcome rating of experiment 2 from all three levels of stimuli, including both conditioning and test trials. In both experiments, this analysis revealed a significant difference between all three levels, showing that we were able to successfully generate three distinct levels of treatment.

In both experiments, we found that people who self-treated and had high treatment expectations experienced greater pain relief during test trials, where the same stimulus intensity (30 VAS) was always presented. We also found that there was no interaction between these factors. Using formal model comparison, we were able to provide strong evidence for a *Likelihood Shift Model* or *Prior Shift Model* (a change in the mean of the likelihood term or a change in the mean of the prior term caused by self-treatment) over the *Likelihood Precision Modulation Model* (modulating the precision of sensory consequences of self-generated outcomes).

The analysis of the EEG data focused on two phases. In the first phase, we examined the EEG power shortly after the presentation of the cue, which signaled agency and treatment effectivity. In the second phase, we analyzed the time frame that included the pain relief phase and the outcome phase.

Modulation of oscillatory activity by differences in stimulus intensity, agency and expectation effects were evident in the EEG data: differences in stimulus intensity in 10 VAS versus 50 VAS conditioning trials were linked to differences in the theta (4-8Hz), alpha-to-beta (8-30Hz), and low gamma (30-50Hz) bands. Agency also

modulated low-frequency oscillatory responses in the alpha-to-beta (8-30Hz) range and low gamma (30-50Hz) range at treatment outcome. A negative cluster of activity in the theta-to-alpha (4-12Hz) range temporally associated with the cue was linked to the interaction of expectation and agency, and was also significantly related to trial-by-trial differences in 30 VAS test trials.

### 6.3.4 Discussion

As expected, our analysis showed a clear difference between 10 VAS and 50 VAS conditioning trials in terms of pain ratings and oscillatory activity. Higher treatment success by an actual larger decrease of stimulus intensity was associated with decreased alpha-to-beta (8-30Hz) and theta (4-8Hz) activity during the relief phase. In the outcome phase, higher treatment success was associated with increased alpha-to-beta and decreased theta activity. This is consistent with previous findings that associate lower alpha-to-beta power (8-30Hz) and higher theta power (4-8Hz) with higher stimulus intensity (Ploner et al., 2017). Our study extends these findings to a treatment context, where higher alpha-to-beta power (8-30Hz) and lower theta power (4-8Hz) were associated with more successful treatment during the outcome phase. This is complementary to results of oscillatory activity in pain processing (see 5.0), where decreases of alpha activity and increases of theta activity is related to stimulus intensity – here, we show that successful treatment leads to a reversal of this pattern: Higher treatment success is associated with a synchronization of alpha-to-beta activity and a desynchronization of theta activity.

We observed modulations by stimulus intensity and agency in outcome-locked EEG data, but not by expectations. This is consistent with previous studies that have shown cue-related expectation effects in the alpha-to-beta band before painful stimulation, but not during painful stimulation (Nickel et al., 2022; Strube et al., 2021a). In another study, pain-induced alpha and gamma responses were significantly influenced by stimulus intensity but not by placebo hypoalgesia (Tiemann et al., 2015). This indicates that expectations are associated with cue-locked effects and may be encoded in oscillatory processes of brain regions that are commonly involved in contextual top-down processing. In contrast, agency had an effect on both cue-locked activity (as an interaction) and outcome-locked activity (as a main effect), indicating influences on both expectations and sensory processing.

In this experiment, we investigated the mechanism behind improved treatment effectivity when the treatment was self-initiated. The *Forward Model* proposes that small prediction errors during self-generated movement result in a perception of a less intense sensation compared to externally generated unpredicted outcomes (Blakemore et al., 1998, 2000). In the context of our pain protocol, this suggests that self-treatment always leads to improved treatment outcomes. In contrast, *Active Inference* suggests that decreased precision of sensory consequences underlies the sensory attenuation phenomenon (Brown et al., 2013). From a Bayesian perspective, pain perception can be seen as the integration of expectation (prior) and nociceptive input (likelihood), with the precision of each term determining its contribution. Therefore, in *Active Inference*, reduced nociceptive precision during our experiment should lead to an increase in the influence of expectation, as its precision remains constant. As a result, we expected that the relative influence of prior expectations would be increased compared to sensory evidence, which would be attenuated in precision. This means that self-treatment should lead to a greater influence of treatment expectation compared to external treatment. From a statistical perspective, this would manifest as an interaction between agency and expectation, i.e. larger differences between low and high treatment expectations in self-treatment compared to external treatment.

Our data showed clear effects of sensory attenuation and treatment expectations in two experiments with different pain rating modalities. However, our data did not show a significant interaction between agency and treatment expectation effects, which strongly favors the *Forward Model* of perception for self-initiated pain treatment. This is supported by Bayesian model comparison using VBA (Daunizeau et al., 2014), which strongly favored the *Likelihood Shift model* or the *Prior Shift Model* over the *Likelihood Precision Modulation Model*. EEG data showed that there was a negative cluster of activity in the theta-to-alpha range (4-12Hz) that was correlated with trial-by-trial VAS ratings in test trials. The early onset of the cluster, which was associated with cue onset, suggests that the modulation of expectations by agency affects the prior rather than the likelihood term, and thus supports the *Prior Shift Model* over the *Likelihood Shift Model*.

The treatment of pain is enhanced by both agency and positive expectations. The sensory attenuation and objectively different stimulus intensities modify the oscillatory

activity during the relief and outcome phases of pain treatment. Expectation effects are associated with EEG activity directly following the cue and interact with agency. Bayesian model comparisons of our data did not find any evidence of a decrease in likelihood precision during self-treatment, which suggests that the positive effect of self-treatment is due to a mean shift as the underlying mechanism in the *Bayesian Pain Model*.

In this study, we sought to understand how the contextual aspect of agency can be integrated within the framework of a *Bayesian Pain Model* and whether this supports the predictions of *Active Inference*. Using EEG data, we determined the oscillatory mechanisms associated with pain reduction during treatment and showed that expectations and agency modulate these signals. Our findings suggest that agency can modulate the effectivity of treatment and that this effect can be explained as a shift in intensity expectations. Furthermore, our data indicate that the modulation of likelihood precision, as derived from *Active Inference*, is not able to explain this effect. Overall, our findings provide insight into the neural mechanisms underlying the placebo/nocebo effect and the role of agency in pain perception.

In the context of *Active Inference*, this is a significant result: From the considerations on *Active Inference*, we deduced that agency should lead to a modulation of the likelihood precision. We found no evidence for this in this study. Thus, we question whether the proposed mechanism of reducing the precision of self-produced action outcomes is indeed causal for sensory attenuation effects and whether this mechanism is suitable for enabling action.

## 7.0 Conclusion

In this cumulative dissertation, we have interpreted the brain as a statistical machine that performs Bayes-optimal inferences about the world. This view is prominent in many contemporary theories of the brain, and evident at both, high-level and low-level systems of the brain.

The descending and ascending recurrent pain network is suited to perform those computations and to relate top-down contextual factors such as expectations and agency to sensory information to enable inference about a pain signal. This ultimately leads to the *Bayesian Pain Model*, which postulates an integration of prior information and nociceptive sensory input to form a pain percept.

While several studies have identified candidate areas for *Predictive Coding* computations, there has been a lack of information about the temporal and spectral orchestration of these signals, which has been identified in the EEG results of study 1 (Strube et al., 2021a; Reprint 15.1). In our sample, expectations were encoded in low frequency bands, while absolute prediction errors were encoded in higher frequencies. In a control condition representing the identical experimental procedure with affective image stimuli representing the visual modality, a fundamentally different pattern emerged, suggesting modality-specific encoding of these signals.

However, there are further contextual factors beyond expectancy effects that modulate pain experience and thus need to be integrated in a *Bayesian Pain Model*. In this work, we addressed agency in pain experience and derived hypotheses from Karl Friston's *Free Energy Principle* and *Active Inference* and tested them against the classic *Forward Model* by integrating predictions from these theories into the *Bayesian Pain Model*. We showed that agency can be integrated into the *Bayesian Pain Model* using a shift in expectations, resulting in pain treatment being more effective.

Based on the principles of *Active Inference*, we expected that agency would lead to a modulation of likelihood precision. However, our study did not find any evidence to support this idea. Therefore, we question whether the proposed mechanism of reducing the precision of self-produced action outcomes is actually responsible for sensory attenuation effects, and whether this mechanism is sufficient to explain enabling of action. Namely, it is postulated that an attenuation of the precision of sensory information in self-produced action outcomes is a necessary condition for action to occur. The minimization of prediction errors can be enabled via two paths: First, an update of the model of the world can occur. Second, an action can be used to adapt the world to the models; for this path, proprioceptive prediction errors must be suppressed. However, we did not find evidence for this mechanism in our data.

Overall, this work highlights important elements that contribute to the understanding of the *Bayesian Brain* in pain.

1) The brain processes nociceptive stimuli via *Predictive Coding* mechanisms. It utilizes recurrent transmissions of top-down expectations and bottom-up prediction errors. These processes are reflected in modality-specific oscillatory patterns.

2) A pain percept is generated in a Bayes-optimal manner as explained by *a Bayesian Pain Mode*l, as has been shown by bi-directional modulations of the pain percept by placebo and nocebo expectations.

3) The *Bayesian Pain Model* can explain contextual modulations such as a reduction in pain by agency via a shift in the mean of the prior, representing intensity expectations. This is in contradiction to ideas of *Active Inference*, which posits that there should be an attenuation of sensory precision by agency.

In conclusion, this dissertation has focused on the application of Bayesian principles to the interpretation of the brain as a statistical machine that performs optimal inferences about the world, with a specific emphasis on pain processing. Our findings show that the brain processes nociceptive stimuli using *Predictive Coding* mechanisms, reflected in modality-specific oscillatory patterns, to generate a Bayes-optimal pain percept. However, our results do not support the idea that certain contextual factors, such as agency, modulate the precision of sensory information in the way previously proposed by theories of *Active Inference*. This calls into question the sufficiency of this mechanism in explaining the enabling of action in the context of pain.

## 8.0 References


Albu, S., & Meagher, M. W. (2016). Expectation of nocebo hyperalgesia affects EEG alpha-activity. *International Journal of Psychophysiology*, *109*, 147–152. https://doi.org/10.1016/j.ijpsycho.2016.08.009

Anchisi, D., & Zanon, M. (2015). A Bayesian Perspective on Sensory and Cognitive Integration in Pain Perception and Placebo Analgesia. *PLOS ONE*, *10*(2), e0117270. https://doi.org/10.1371/journal.pone.0117270

Andersen, E., & Dafny, N. (1983). An ascending serotonergic pain modulation pathway from the dorsal raphe nucleus to the parafascicularis nucleus of the thalamus. *Brain Research*, *269*(1), 57–67. https://doi.org/10.1016/0006-8993(83)90962-9

Apkarian, A. V., Bushnell, M. C., Treede, R.-D., & Zubieta, J.-K. (2005). Human brain mechanisms of pain perception and regulation in health and disease. *European Journal of Pain (London, England)*, *9*(4), 463–484. https://doi.org/10.1016/j.ejpain.2004.11.001

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*, 390–398. https://doi.org/10.1016/j.tics.2012.05.003

Atlas, L. Y., & Wager, T. D. (2014). A meta-analysis of brain mechanisms of placebo analgesia: Consistent findings and unanswered questions. *Handbook of Experimental Pharmacology*, *225*, 37–69. https://doi.org/10.1007/978-3-662-44519-8_3

Babiloni, C., Brancucci, A., Babiloni, F., Capotosto, P., Carducci, F., Cincotti, F., Arendt-Nielsen, L., Chen, A. C. N., & Rossini, P. M. (2003). Anticipatory cortical responses during the expectancy of a predictable painful stimulation. A high-resolution electroencephalography study. *European Journal of*

*Neuroscience*, *18*(6), 1692–1700. https://doi.org/10.1046/j.1460-9568.2003.02851.x

Ballantyne, J. C., Carr, D. B., Chalmers, T. C., Dear, K. B., Angelillo, I. F., & Mosteller, F. (1993). Postoperative patient-controlled analgesia: Meta-analyses of initial randomized control trials. *Journal of Clinical Anesthesia*, *5*(3), 182–193. https://doi.org/10.1016/0952-8180(93)90013-5

Bantick, S. J., Wise, R. G., Ploghaus, A., Clare, S., Smith, S. M., & Tracey, I. (2002). Imaging how attention modulates pain in humans using functional MRI. *Brain*, *125*(2), 310–319. https://doi.org/10.1093/brain/awf022

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1–23. https://doi.org/10.1093/scan/nsw154

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews. Neuroscience*, *16*(7), 419–429. https://doi.org/10.1038/nrn3950

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*(4), 695–711. https://doi.org/10.1016/j.neuron.2012.10.038

Bauer, M., Stenner, M.-P., Friston, K. J., & Dolan, R. J. (2014). Attentional modulation of alpha/beta and gamma oscillations reflect functionally distinct processes. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(48), 16117–16125. https://doi.org/10.1523/JNEUROSCI.3474-13.2014

Beck, B., Di Costa, S., & Haggard, P. (2017). Having control over the external world increases the implicit sense of agency. *Cognition*, *162*, 54–60. https://doi.org/10.1016/j.cognition.2017.02.002

Bessou, P., & Perl, E. R. (1969). Response of cutaneous sensory units with unmyelinated fibers to noxious stimuli. *Journal of Neurophysiology*, *32*(6), 1025–1043. https://doi.org/10.1152/jn.1969.32.6.1025

Bingel, U., Lorenz, J., Schoell, E., Weiller, C., & Büchel, C. (2006). Mechanisms of placebo analgesia: RACC recruitment of a subcortical antinociceptive network. *Pain*, *120*(1–2), 8–15. https://doi.org/10.1016/j.pain.2005.08.027

Blakemore, S.-J., Frith, C. D., & Wolpert, D. M. (1999). Spatio-Temporal Prediction Modulates the Perception of Self-Produced Stimuli. *Journal of Cognitive Neuroscience*, *11*(5), 551–559. https://doi.org/10.1162/089892999563607

Blakemore, S.-J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself? *NeuroReport*, *11*(11), R11.

Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience*, *1*(7), Art. 7. https://doi.org/10.1038/2870

Boucher, O., D'Hondt, F., Tremblay, J., Lepore, F., Lassonde, M., Vannasing, P., Bouthillier, A., & Nguyen, D. K. (2015). Spatiotemporal dynamics of affective picture processing revealed by intracranial high-gamma modulations. *Human Brain Mapping*, *36*(1), 16–28. https://doi.org/10.1002/hbm.22609

Bradley, M. M., & Lang, P. J. (2007). The International Affective Picture System (IAPS) in the study of emotion and attention. In *Handbook of emotion elicitation and assessment* (S. 29–46). Oxford University Press.

Bromm, B., & Treede, R.-D. (1980). Withdrawal reflex, skin resistance reaction and pain ratings due to electrical stimuli in man. *PAIN*, *9*(3), 339–354. https://doi.org/10.1016/0304-3959(80)90048-2

Brown, H., Adams, R. A., Parees, I., Edwards, M., & Friston, K. (2013). Active inference, sensory attenuation and illusions. *Cognitive Processing*, *14*(4), 411–427. https://doi.org/10.1007/s10339-013-0571-3

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

Buffalo, E. A., Fries, P., Landman, R., Buschman, T. J., & Desimone, R. (2011). Laminar differences in gamma and alpha coherence in the ventral stream. *Proceedings of the National Academy of Sciences*, *108*(27), 11262–11267. https://doi.org/10.1073/pnas.1011284108

Burgess, P. R., & Perl, E. R. (1967). Myelinated afferent fibres responding specifically to noxious stimulation of the skin. *The Journal of Physiology*, *190*(3), 541–562. https://doi.org/10.1113/jphysiol.1967.sp008227

Chen, A. C., & Rappelsberger, P. (1994). Brain and human pain: Topographic EEG amplitude and coherence mapping. *Brain Topography*, *7*(2), 129–140. https://doi.org/10.1007/BF01186771

Claxton, G. (1975). Why can't we tickle ourselves? *Perceptual and Motor Skills*, *41*(1), 335–338. https://doi.org/10.2466/pms.1975.41.1.335

Colloca, L., & Benedetti, F. (2005). Placebos and painkillers: Is mind as real as matter? *Nature Reviews. Neuroscience*, *6*(7), 545–552. https://doi.org/10.1038/nrn1705

Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nature Reviews. Neuroscience*, *3*(8), 655–666. https://doi.org/10.1038/nrn894

Cui, Y., Versace, F., Engelmann, J. M., Minnix, J. A., Robinson, J. D., Lam, C. Y., Karam-Hage, M., Brown, V. L., Wetter, D. W., Dani, J. A., Kosten, T. R., &

Cinciripini, P. M. (2013). Alpha oscillations in response to affective and cigarette-related stimuli in smokers. *Nicotine & Tobacco Research: Official Journal of the Society for Research on Nicotine and Tobacco*, *15*(5), 917–924. https://doi.org/10.1093/ntr/nts209

Darwin, C. (1872). *The Expression of Emotions in Man and Animals*. John Murray.

Daunizeau, J., Adam, V., & Rigoux, L. (2014). VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data. *PLOS Computational Biology*, *10*(1), e1003441. https://doi.org/10.1371/journal.pcbi.1003441

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The Helmholtz machine. *Neural Computation*, *7*(5), 889–904. https://doi.org/10.1162/neco.1995.7.5.889

De Cesarei, A., & Codispoti, M. (2011). Affective modulation of the LPP and α-ERD during picture viewing. *Psychophysiology*, *48*(10), 1397–1404. https://doi.org/10.1111/j.1469-8986.2011.01204.x

Del Percio, C., Le Pera, D., Arendt-Nielsen, L., Babiloni, C., Brancucci, A., Chen, A. C. N., De Armas, L., Miliucci, R., Restuccia, D., Valeriani, M., & Rossini, P. M. (2006). Distraction affects frontal alpha rhythms related to expectancy of pain: An EEG study. *NeuroImage*, *31*(3), 1268–1277. https://doi.org/10.1016/j.neuroimage.2006.01.013

Dubin, A. E., & Patapoutian, A. (2010). Nociceptors: The sensors of the pain pathway. *The Journal of Clinical Investigation*, *120*(11), 3760–3772. https://doi.org/10.1172/JCI42843

Dum, R. P., Levinthal, D. J., & Strick, P. L. (2009). The spinothalamic system targets motor and sensory areas in the cerebral cortex of monkeys. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(45), 14223–14235. https://doi.org/10.1523/JNEUROSCI.3398-09.2009

Edwards, E., Soltani, M., Deouell, L. Y., Berger, M. S., & Knight, R. T. (2005). High gamma activity in response to deviant auditory stimuli recorded directly from human cortex. *Journal of Neurophysiology*, *94*(6), 4269–4280. https://doi.org/10.1152/jn.00324.2005

Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and Surprise Determine Neural Population Responses in the Ventral Visual Stream. *Journal of Neuroscience*, *30*(49), 16601–16608. https://doi.org/10.1523/JNEUROSCI.2770-10.2010

Eippert, F., Finsterbusch, J., Bingel, U., & Büchel, C. (2009). Direct evidence for spinal cord involvement in placebo analgesia. *Science (New York, N.Y.)*, *326*(5951), 404. https://doi.org/10.1126/science.1180142

Fazeli, S., & Büchel, C. (2018). Pain-Related Expectation and Prediction Error Signals in the Anterior Insula Are Not Related to Aversiveness. *The Journal of Neuroscience*, *38*(29), 6461. https://doi.org/10.1523/JNEUROSCI.0671-18.2018

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex (New York, N.Y.*, *1*(1), 1–47. https://doi.org/10.1093/cercor/1.1.1-a

Ferracuti, S., Seri, S., Mattia, D., & Cruccu, G. (1994). Quantitative EEG modifications during the Cold Water Pressor Test: Hemispheric and hand differences. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *17*(3), 261–268. https://doi.org/10.1016/0167-8760(94)90068-x

Fields, H. L., & Heinricher, M. M. (1989). Brainstem modulation of nociceptor-driven withdrawal reflexes. *Annals of the New York Academy of Sciences*, *563*, 34–44. https://doi.org/10.1111/j.1749-6632.1989.tb42188.x

Franciotti, R., Ciancetta, L., Della Penna, S., Belardinelli, P., Pizzella, V., & Romani, G. L. (2009). Modulation of alpha oscillations in insular cortex reflects the threat of painful stimuli. *NeuroImage*, *46*(4), 1082–1090. https://doi.org/10.1016/j.neuroimage.2009.03.034

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K. (2012). The history of the future of the Bayesian brain. *NeuroImage*, *62*(2), 1230–1233. https://doi.org/10.1016/j.neuroimage.2011.10.004

Friston, K., Daunizeau, J., & Kiebel, S. J. (2009). Reinforcement Learning or Active Inference? *PLOS ONE*, *4*(7), e6421. https://doi.org/10.1371/journal.pone.0006421

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active Inference: A Process Theory. *Neural Computation*, *29*(1), 1–49. https://doi.org/10.1162/NECO_a_00912

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1211–1221. https://doi.org/10.1098/rstb.2008.0300

Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology-Paris*, *100*(1), 70–87. https://doi.org/10.1016/j.jphysparis.2006.10.001

Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., & Penny, W. (2007). Variational free energy and the Laplace approximation. *NeuroImage*, *34*(1), 220–234. https://doi.org/10.1016/j.neuroimage.2006.08.035

Friston, K., & Penny, W. (2011). Post hoc Bayesian model selection. *NeuroImage*, *56*(4), 2089–2099. https://doi.org/10.1016/j.neuroimage.2011.03.062

Friston, K., Schwartenbeck, P., Fitzgerald, T., Moutoussis, M., Behrens, T., & Dolan, R. (2013). The anatomy of choice: Active inference and agency. *Frontiers in Human Neuroscience*, *7*. https://www.frontiersin.org/article/10.3389/fnhum.2013.00598

Geuter, S., Boll, S., Eippert, F., & Büchel, C. (2017). Functional dissociation of stimulus intensity encoding and predictive coding of pain in the insula. *eLife*, *6*, e24770. https://doi.org/10.7554/eLife.24770

Geuter, S., & Buchel, C. (2013). Facilitation of Pain in the Human Spinal Cord by Nocebo Treatment. *Journal of Neuroscience*, *33*(34), 13784–13790. https://doi.org/10.1523/JNEUROSCI.2191-13.2013

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

Gniß, S., Kappesser, J., & Hermann, C. (2020). Placebo effect in children: The role of expectation and learning. *PAIN*, *161*(6), 1191–1201. https://doi.org/10.1097/j.pain.0000000000001811

Grahl, A., Onat, S., & Büchel, C. (2018). The periaqueductal gray and Bayesian integration in placebo analgesia. *eLife*, *7*, e32930. https://doi.org/10.7554/eLife.32930

Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *290*(1038), 181–197. https://doi.org/10.1098/rstb.1980.0090

Gregory, R. L., Young, J. Z., & Longuet-Higgins, H. C. (1968). Perceptual illusions and brain models. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *171*(1024), 279–296. https://doi.org/10.1098/rspb.1968.0071

Güntekin, B., & Tülay, E. (2014). Event related beta and gamma oscillatory responses during perception of affective pictures. *Brain Research*, *1577*, 45–56. https://doi.org/10.1016/j.brainres.2014.06.029

Hald, L. A., Bastiaansen, M. C. M., & Hagoort, P. (2006). EEG theta and gamma responses to semantic violations in online sentence processing. *Brain and Language*, *96*(1), 90–105. https://doi.org/10.1016/j.bandl.2005.06.007

Halligan, P. W. (2002). Phantom limbs: the body in mind. Cognitive Neuropsychiatry,7(3), 251-269.

Hammond, D. (1989). New Insights Regarding Organization of Spinal Cord Pain Pathways. *Physiology*, *4*(3), 98–101. https://doi.org/10.1152/physiologyonline.1989.4.3.98

Hauck, M., Lorenz, J., & Engel, A. K. (2007). Attention to painful stimulation enhances gamma-band activity and synchronization in human sensorimotor cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *27*(35), 9270–9277. https://doi.org/10.1523/JNEUROSCI.2283-07.2007

Hauck, M., Metzner, S., Rohlffs, F., Lorenz, J., & Engel, A. K. (2013). The influence of music and music therapy on pain-induced neuronal oscillations measured by magnetencephalography. *PAIN*, *154*(4), 539–547. https://doi.org/10.1016/j.pain.2012.12.016

Heilbron, M., & Chait, M. (2018). Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? *Neuroscience*, *389*, 54–73. https://doi.org/10.1016/j.neuroscience.2017.07.061

Helmchen, C., Mohr, C., Erdmann, C., Binkofski, F., & Büchel, C. (2006). Neural activity related to self- versus externally generated painful stimuli reveals

distinct differences in the lateral pain system in a parametric fMRI study. *Human brain mapping*, *27*, 755–765. https://doi.org/10.1002/hbm.20217

Helmholtz, H. von. (1867). *Handbuch der physiologischen Optik*. Voss.

Herrmann, C. S., Lenz, D., Junge, S., Busch, N. A., & Maess, B. (2004). Memory-matches evoke human gamma-responses. *BMC Neuroscience*, *5*, 13. https://doi.org/10.1186/1471-2202-5-13

Herrmann, C. S., Munk, M. H. J., & Engel, A. K. (2004). Cognitive functions of gamma-band activity: Memory match and utilization. *Trends in Cognitive Sciences*, *8*(8), 347–355. https://doi.org/10.1016/j.tics.2004.06.006

Hesselmann, G., Sadaghiani, S., Friston, K. J., & Kleinschmidt, A. (2010). Predictive Coding or Evidence Accumulation? False Inference and Neuronal Fluctuations. *PLOS ONE*, *5*(3), e9926. https://doi.org/10.1371/journal.pone.0009926

Hird, E. J., Jones, A. K. P., Talmi, D., & El-Deredy, W. (2018). A comparison between the neural correlates of laser and electric pain stimulation and their modulation by expectation. *Journal of Neuroscience Methods*, *293*, 117–127. https://doi.org/10.1016/j.jneumeth.2017.09.011

Horing, B., & Büchel, C. (2022). The human insula processes both modality-independent and pain-selective learning signals. *PLOS Biology*, *20*(5), e3001540. https://doi.org/10.1371/journal.pbio.3001540

Horing, B., Sprenger, C., & Büchel, C. (2019). The parietal operculum preferentially encodes heat pain and not salience. *PLoS Biology*, *17*(8), e3000205. https://doi.org/10.1371/journal.pbio.3000205

Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *WIREs Cognitive Science*, *2*(5), 580–593. https://doi.org/10.1002/wcs.142

Huneke, N. T. M., Brown, C. A., Burford, E., Watson, A., Trujillo-Barreto, N. J., El-Deredy, W., & Jones, A. K. P. (2013). Experimental Placebo Analgesia Changes Resting-State Alpha Oscillations. *PLoS ONE*, *8*(10), e78278. https://doi.org/10.1371/journal.pone.0078278

Jehee, J. F. M., & Ballard, D. H. (2009). Predictive Feedback Can Account for Biphasic Responses in the Lateral Geniculate Nucleus. *PLOS Computational Biology*, *5*(5), e1000373. https://doi.org/10.1371/journal.pcbi.1000373

Jensen, M. P., & Karoly, P. (1991). Control beliefs, coping efforts, and adjustment to chronic pain. *Journal of Consulting and Clinical Psychology*, *59*(3), 431–438. https://doi.org/10.1037//0022-006x.59.3.431

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*(2), 163–178.

Kakigi, R., & Shibasaki, H. (1992). Mechanisms of pain relief by vibration and movement. *Journal of Neurology, Neurosurgery & Psychiatry*, *55*(4), 282–286. https://doi.org/10.1136/jnnp.55.4.282

Kaplan, R., & Friston, K. J. (2018). Planning and navigation as active inference. *Biological Cybernetics*, *112*(4), 323–343. https://doi.org/10.1007/s00422-018-0753-2

Karsh, N., Goldstein, O., & Eitam, B. (2018). Evidence for pain attenuation by the motor system-based judgment of agency. *Consciousness and Cognition*, *57*, 134–146. https://doi.org/10.1016/j.concog.2017.11.012

Keil, A., Müller, M. M., Gruber, T., Wienbruch, C., Stolarova, M., & Elbert, T. (2001). Effects of emotional arousal in the cerebral hemispheres: A study of oscillatory brain activity and event-related potentials. *Clinical Neurophysiology: Official*

*Journal of the International Federation of Clinical Neurophysiology*, *112*(11), 2057–2068. https://doi.org/10.1016/s1388-2457(01)00654-x

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, *55*, 271–304. https://doi.org/10.1146/annurev.psych.55.090902.142005

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. https://doi.org/10.1016/j.tins.2004.10.007

LaMotte, R. H., & Campbell, J. N. (1978). Comparison of responses of warm and nociceptive C-fiber afferents in monkey with human judgments of thermal pain. *Journal of Neurophysiology*, *41*(2), 509–528. https://doi.org/10.1152/jn.1978.41.2.509

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *20*(7), 1434–1448. https://doi.org/10.1364/josaa.20.001434

Lesicko, A. M., Angeloni, C. F., Blackwell, J. M., De Biasi, M., & Geffen, M. N. (2022). Corticofugal regulation of predictive coding. *eLife*, *11*, e73289. https://doi.org/10.7554/eLife.73289

Liu, C.-C., Ohara, S., Franaszczuk, P. J., & Lenz, F. A. (2011). Attention to painful cutaneous laser stimuli evokes directed functional connectivity between activity recorded directly from human pain-related cortical structures. *Pain*, *152*(3), 664–675. https://doi.org/10.1016/j.pain.2010.12.016

Lorenz, J., Hauck, M., Paur, R. C., Nakamura, Y., Zimmermann, R., Bromm, B., & Engel, A. K. (2005). Cortical correlates of false expectations during pain intensity judgments—A possible manifestation of placebo/nocebo cognitions.

*Brain, Behavior, and Immunity*, *19*(4), 283–295. https://doi.org/10.1016/j.bbi.2005.03.010

Maier, A., Adams, G., Aura, C., & Leopold, D. (2010). Distinct Superficial and Deep Laminar Domains of Activity in the Visual Cortex during Rest and Stimulation. *Frontiers in Systems Neuroscience*, *4*. https://www.frontiersin.org/articles/10.3389/fnsys.2010.00031

Makeig, S., Bell, A., Jung, T.-P., & Sejnowski, T. (1996). *Independent Component Analysis of Electroencephalographic Data*. *8*.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Martini, N., Menicucci, D., Sebastiani, L., Bedini, R., Pingitore, A., Vanello, N., Milanesi, M., Landini, L., & Gemignani, A. (2012). The dynamics of EEG gamma responses to unpleasant visual stimuli: From local activity to functional connectivity. *NeuroImage*, *60*(2), 922–932. https://doi.org/10.1016/j.neuroimage.2012.01.060

Maunsell, J. H., & van Essen, D. C. (1983). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *3*(12), 2563–2586. https://doi.org/10.1523/JNEUROSCI.03-12-02563.1983

May, A. (2008). Chronic pain may change the structure of the brain. *PAIN*, *137*(1), 7–15. https://doi.org/10.1016/j.pain.2008.02.034

May, A. (2011). Structural Brain Imaging: A Window into Chronic Pain. *The Neuroscientist*, *17*(2), 209–220. https://doi.org/10.1177/1073858410396220

May, E. S., Butz, M., Kahlbrock, N., Hoogenboom, N., Brenner, M., & Schnitzler, A. (2012). Pre- and post-stimulus alpha activity shows differential modulation with spatial attention during the processing of pain. *NeuroImage*, *62*(3), 1965–1974. https://doi.org/10.1016/j.neuroimage.2012.05.071

McNicol, E. D., Ferguson, M. C., & Hudcova, J. (2015). Patient controlled opioid analgesia versus non-patient controlled opioid analgesia for postoperative pain. *Cochrane Database of Systematic Reviews*, *6*. https://doi.org/10.1002/14651858.CD003348.pub3

Miron, D., Duncan, G. H., & Catherine Bushnell, M. (1989). Effects of attention on the intensity and unpleasantness of thermal pain. *Pain*, *39*(3), 345–352. https://doi.org/10.1016/0304-3959(89)90048-1

Mohr, C., Leyendecker, S., & Helmchen, C. (2008). Dissociable neural activity to self- vs. externally administered thermal hyperalgesia: A parametric fMRI study. *European Journal of Neuroscience*, *27*(3), 739–749. https://doi.org/10.1111/j.1460-9568.2008.06036.x

Mohr, C., Leyendecker, S., Petersen, D., & Helmchen, C. (2012). Effects of perceived and exerted pain control on neural activity during pain relief in experimental heat hyperalgesia: A fMRI study. *European Journal of Pain (London, England)*, *16*(4), 496–508. https://doi.org/10.1016/j.ejpain.2011.07.010

Molinaro, N., Barraza, P., & Carreiras, M. (2013). Long-range neural synchronization supports fast and efficient reading: EEG correlates of processing expected words in sentences. *NeuroImage*, *72*, 120–132. https://doi.org/10.1016/j.neuroimage.2013.01.031

Moran, R. J., Campo, P., Symmonds, M., Stephan, K. E., Dolan, R. J., & Friston, K. J. (2013). Free energy, precision and learning: The role of cholinergic neuromodulation. *The Journal of Neuroscience: The Official Journal of the*

*Society for Neuroscience*, *33*(19), 8227–8236. https://doi.org/10.1523/JNEUROSCI.4255-12.2013

Müller, M. M., Keil, A., Gruber, T., & Elbert, T. (1999). Processing of affective pictures modulates right-hemispheric gamma band EEG activity. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *110*(11), 1913–1920. https://doi.org/10.1016/s1388-2457(99)00151-0

Müller, M., & Netter, P. (2000). Relationship of subjective helplessness and pain perception after electric skin stimuli. *Stress Medicine*, *16*(2), 109–115. https://doi.org/10.1002/(SICI)1099-1700(200003)16:2<109::AID-SMI837>3.0.CO;2-#

Neziri, A. Y., Andersen, O. K., Petersen-Felix, S., Radanov, B., Dickenson, A. H., Scaramozzino, P., Arendt-Nielsen, L., & Curatolo, M. (2010). The nociceptive withdrawal reflex: Normative values of thresholds and reflex receptive fields. *European Journal of Pain*, *14*(2), 134–141. https://doi.org/10.1016/j.ejpain.2009.04.010

Nickel, M. M., Tiemann, L., Hohn, V. D., May, E. S., Gil Ávila, C., Eippert, F., & Ploner, M. (2022). Temporal–spectral signaling of sensory information and expectations in the cerebral processing of pain. *Proceedings of the National Academy of Sciences*, *119*(1), e2116616119. https://doi.org/10.1073/pnas.2116616119

Obleser, J., & Kotz, S. A. (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *NeuroImage*, *55*(2), 713–723. https://doi.org/10.1016/j.neuroimage.2010.12.020

Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., & Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation:

Functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(13), 5951–5955. https://doi.org/10.1073/pnas.89.13.5951

Ohara, S., Crone, N. E., Weiss, N., & Lenz, F. A. (2006). Analysis of synchrony demonstrates 'pain networks' defined by rapidly switching, task-specific, functional connectivity between pain-related cortical structures. *PAIN*, *123*(3), 244–253. https://doi.org/10.1016/j.pain.2006.02.012

Ongaro, G., & Kaptchuk, T. J. (2019). Symptom perception, placebo effects, and the Bayesian brain. *Pain*, *160*(1), 1–4. https://doi.org/10.1097/j.pain.0000000000001367

Osipova, D., Takashima, A., Oostenveld, R., Fernández, G., Maris, E., & Jensen, O. (2006). Theta and gamma oscillations predict encoding and retrieval of declarative memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *26*(28), 7523–7531. https://doi.org/10.1523/JNEUROSCI.1948-06.2006

Parkes, C. M. (1973). Factors determining the persistence of phantom pain in the amputee. *Journal of Psychosomatic Research*, *17*(2), 97–108. https://doi.org/10.1016/0022-3999(73)90010-X

Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. MIT Press.

Parras, G. G., Nieto-Diego, J., Carbajal, G. V., Valdés-Baizabal, C., Escera, C., & Malmierca, M. S. (2017). Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nature Communications*, *8*(1), 2148. https://doi.org/10.1038/s41467-017-02038-6

Peerdeman, K. J., van Laarhoven, A. I. M., Keij, S. M., Vase, L., Rovers, M. M., Peters, M. L., & Evers, A. W. M. (2016). Relieving patients' pain with

expectation interventions: A meta-analysis. *Pain*, *157*(6), 1179–1191. https://doi.org/10.1097/j.pain.0000000000000540

Pellino, T. A., & Ward, S. E. (1998). Perceived control mediates the relationship between pain severity and patient satisfaction. *Journal of Pain and Symptom Management*, *15*(2), 110–116.

Peng, W., Hu, L., Zhang, Z., & Hu, Y. (2014). Changes of spontaneous oscillatory activity to tonic heat pain. *PloS One*, *9*(3), e91052. https://doi.org/10.1371/journal.pone.0091052

Pervin, L. A. (1963). The need to predict and control under conditions of threat. *Journal of Personality*, *31*(4), 570–587. https://doi.org/10.1111/j.1467-6494.1963.tb01320.x

Petersen, G. L., Finnerup, N. B., Colloca, L., Amanzio, M., Price, D. D., Jensen, T. S., & Vase, L. (2014). The magnitude of nocebo effects in pain: A meta-analysis. *PAIN®*, *155*(8), 1426–1434. https://doi.org/10.1016/j.pain.2014.04.016

Petrovic, P., Kalso, E., Petersson, K. M., & Ingvar, M. (2002). Placebo and Opioid Analgesia—Imaging a Shared Neuronal Network. *Science*, *295*(5560), 1737–1740. https://doi.org/10.1126/science.1067176

Ploner, M., Sorg, C., & Gross, J. (2017). Brain Rhythms of Pain. *Trends in Cognitive Sciences*, *21*(2), 100–110. https://doi.org/10.1016/j.tics.2016.12.001

Qiao, J.-T., & Dafny, N. (1988). Dorsal raphe stimulation modulates nociceptive responses in thalamic parafascicular neurons via an ascending pathway: Further studies on ascending pain modulation pathways. *Pain*, *34*(1), 65–74. https://doi.org/10.1016/0304-3959(88)90183-2

Ramachandran, V. S., & Rogers-Ramachandran, D. (1996). Synaesthesia in phantom limbs induced with mirrors. *Proceedings of the Royal Society of*

*London. Series B: Biological Sciences*, *263*(1369), 377–386. https://doi.org/10.1098/rspb.1996.0058

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), Art. 1. https://doi.org/10.1038/4580

Schenk, L. A., Sprenger, C., Onat, S., Colloca, L., & Büchel, C. (2017). Suppression of Striatal Prediction Errors by the Prefrontal Cortex in Placebo Hypoalgesia. *The Journal of Neuroscience*, *37*(40), 9715–9723. https://doi.org/10.1523/JNEUROSCI.1101-17.2017

Schneider, T. R., Hipp, J. F., Domnick, C., Carl, C., Büchel, C., & Engel, A. K. (2018). Modulation of neuronal oscillatory activity in the beta- and gamma-band is associated with current individual anxiety levels. *NeuroImage*, *178*, 423–434. https://doi.org/10.1016/j.neuroimage.2018.05.059

Schouenborg, J., Holmberg, H., & Weng, H.-R. (1992). Functional organization of the nociceptive withdrawal reflexes. *Experimental Brain Research*, *90*(3), 469–478. https://doi.org/10.1007/BF00230929

Schubring, D., & Schupp, H. T. (2019). Affective picture processing: Alpha- and lower beta-band desynchronization reflects emotional arousal. *Psychophysiology*, *56*(8), e13386. https://doi.org/10.1111/psyp.13386

Schubring, D., & Schupp, H. T. (2021). Emotion and Brain Oscillations: High Arousal is Associated with Decreases in Alpha- and Lower Beta-Band Power. *Cerebral Cortex*, *31*(3), 1597–1608. https://doi.org/10.1093/cercor/bhaa312

Schulz, E., May, E. S., Postorino, M., Tiemann, L., Nickel, M. M., Witkovsky, V., Schmidt, P., Gross, J., & Ploner, M. (2015). Prefrontal Gamma Oscillations Encode Tonic Pain in Humans. *Cerebral Cortex (New York, N.Y.: 1991)*, *25*(11), 4407–4414. https://doi.org/10.1093/cercor/bhv043

Schwartenbeck, P., FitzGerald, T. H. B., Mathys, C., Dolan, R., Wurst, F., Kronbichler, M., & Friston, K. (2015). Optimal inference with suboptimal models: Addiction and active Bayesian inference. *Medical Hypotheses*, *84*(2), 109–117. https://doi.org/10.1016/j.mehy.2014.12.007

Seth, A., Suzuki, K., & Critchley, H. (2012). An Interoceptive Predictive Coding Model of Conscious Presence. *Frontiers in Psychology*, *2*. https://www.frontiersin.org/articles/10.3389/fpsyg.2011.00395

Simmel, M. L. (1959). Phantoms, phantom pain and denial. *American Journal of Psychotherapy*, *13*, 603–613. https://doi.org/10.1176/appi.psychotherapy.1959.13.3.603

Sprenger, C., Eippert, F., Finsterbusch, J., Bingel, U., Rose, M., & Büchel, C. (2012). Attention Modulates Spinal Cord Responses to Pain. *Current Biology*, *22*(11), 1019–1022. https://doi.org/10.1016/j.cub.2012.04.006

Sprenger, C., Finsterbusch, J., & Büchel, C. (2015). Spinal Cord–Midbrain Functional Connectivity Is Related to Perceived Pain Intensity: A Combined Spino-Cortical fMRI Study. *The Journal of Neuroscience*, *35*(10), 4248–4257. https://doi.org/10.1523/JNEUROSCI.4897-14.2015

Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: A fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *216*(1205), 427–459. https://doi.org/10.1098/rspb.1982.0085

Staub, E., Tursky, B., & Schwartz, G. E. (1971). Self-control and predictability: Their effects on reactions to aversive stimulation. *Journal of Personality and Social Psychology*, *18*(2), 157–162. https://doi.org/10.1037/h0030851

Sterzer, P., Frith, C., & Petrovic, P. (2008). Believing is seeing: Expectations alter visual awareness. *Current Biology*, *18*(16), R697–R698. https://doi.org/10.1016/j.cub.2008.06.021

Strube, A., Horing, B., Rose, M., & Büchel, C. (2022). Agency Affects Pain Inference through Intensity Shift as Opposed to Precision Modulation [Manuscript submitted for publication]. Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf

Strube, A., Rose, M., Fazeli, S., & Büchel, C. (2021a). The temporal and spectral characteristics of expectations and prediction errors in pain and thermoception. *eLife*, *10*, e62809. https://doi.org/10.7554/eLife.62809

Strube, A., Rose, M., Fazeli, S., & Büchel, C. (2021b). Alpha-to-beta- and gamma-band activity reflect predictive coding in affective visual processing. *Scientific Reports*, *11*(1), Art. 1. https://doi.org/10.1038/s41598-021-02939-z

Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., & Hirsch, J. (2006). Predictive codes for forthcoming perception in the frontal cortex. *Science (New York, N.Y.)*, *314*(5803), 1311–1314. https://doi.org/10.1126/science.1132028

Taesler, P., & Rose, M. (2016). Prestimulus Theta Oscillations and Connectivity Modulate Pain Perception. *Journal of Neuroscience*, *36*(18), 5026–5033. https://doi.org/10.1523/JNEUROSCI.3325-15.2016

Thomaidou, M. A., Blythe, J. S., Houtman, S. J., Veldhuijzen, D. S., van Laarhoven, A. I. M., & Evers, A. W. M. (2021). Temporal structure of brain oscillations predicts learned nocebo responses to pain. *Scientific Reports*, *11*(1), Art. 1. https://doi.org/10.1038/s41598-021-89368-0

Thomaidou, M. A., Veldhuijzen, D. S., Meulders, A., & Evers, A. W. M. (2021). An experimental investigation into the mediating role of pain-related fear in

boosting nocebo hyperalgesia. *Pain*, *162*(1), 287–299. https://doi.org/10.1097/j.pain.0000000000002017

Thompson, S. C. (1981). Will it hurt less if I can control it? A complex answer to a simple question. *Psychological Bulletin*, *90*(1), 89–101. https://doi.org/10.1037/0033-2909.90.1.89

Tiemann, L., May, E. S., Postorino, M., Schulz, E., Nickel, M. M., Bingel, U., & Ploner, M. (2015). Differential neurophysiological correlates of bottom-up and top-down modulations of pain. *PAIN*, *156*(2), 289–296. https://doi.org/10.1097/01.j.pain.0000460309.94442.44

Tinnermann, A., Geuter, S., Sprenger, C., Finsterbusch, J., & Büchel, C. (2017). Interactions between brain and spinal cord mediate value effects in nocebo hyperalgesia. *Science (New York, N.Y.)*, *358*(6359), 105–108. https://doi.org/10.1126/science.aan1221

Todorovic, A., van Ede, F., Maris, E., & de Lange, F. P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: An MEG study. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(25), 9118–9123. https://doi.org/10.1523/JNEUROSCI.1425-11.2011

Torebjörk, E. (1985). Nociceptor activation and pain. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *308*(1136), 227–234. https://doi.org/10.1098/rstb.1985.0023

Tracey, I., & Mantyh, P. W. (2007). The cerebral signature for pain perception and its modulation. *Neuron*, *55*(3), 377–391. https://doi.org/10.1016/j.neuron.2007.07.012

Uran, C., Peter, A., Lazar, A., Barnes, W., Klon-Lipok, J., Shapcott, K. A., Roese, R., Fries, P., Singer, W., & Vinck, M. (2022). Predictive coding of natural images

by V1 firing rates and rhythmic synchronization. *Neuron*, *110*(7), 1240-1257.e8. https://doi.org/10.1016/j.neuron.2022.01.002

van de Sand, M. F., Menz, M. M., Sprenger, C., & Büchel, C. (2018). Nocebo-induced modulation of cerebral itch processing – An fMRI study. *NeuroImage*, *166*, 209–218. https://doi.org/10.1016/j.neuroimage.2017.10.056

van Pelt, S., Heil, L., Kwisthout, J., Ondobaka, S., van Rooij, I., & Bekkering, H. (2016). Beta- and gamma-band activity reflect predictive coding in the processing of causal events. *Social Cognitive and Affective Neuroscience*, *11*(6), 973–980. https://doi.org/10.1093/scan/nsw017

Wager, T. D., Rilling, J. K., Smith, E. E., Sokolik, A., Casey, K. L., Davidson, R. J., Kosslyn, S. M., Rose, R. M., & Cohen, J. D. (2004). Placebo-induced changes in FMRI in the anticipation and experience of pain. *Science (New York, N.Y.)*, *303*(5661), 1162–1167. https://doi.org/10.1126/science.1093065

Wang, L., Hagoort, P., & Jensen, O. (2018). Gamma Oscillatory Activity Related to Language Prediction. *Journal of Cognitive Neuroscience*, *30*(8), 1075–1085. https://doi.org/10.1162/jocn_a_01275

Wang, Y., Wang, J.-Y., & Luo, F. (2011). Why Self-Induced Pain Feels Less Painful than Externally Generated Pain: Distinct Brain Activation Patterns in Self- and Externally Generated Pain. *PLOS ONE*, *6*(8), e23536. https://doi.org/10.1371/journal.pone.0023536

Weisenberg, M., Wolf, Y., Mittwoch, T., Mikulincer, M., & Aviram, O. (1985). Subject versus experimenter control in the reaction to pain. *Pain*, *23*(2), 187–200. https://doi.org/10.1016/0304-3959(85)90059-4

Weiskrantz, L., Elliott, J., & Darlington, C. (1971). Preliminary Observations on Tickling Oneself. *Nature*, *230*(5296), Art. 5296. https://doi.org/10.1038/230598a0

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

Wiech, K., Jbabdi, S., Lin, C. S., Andersson, J., & Tracey, I. (2014). Differential structural and resting state connectivity between insular subdivisions and other pain-related brain regions. *Pain*, *155*(10), 2047–2055. https://doi.org/10.1016/j.pain.2014.07.009

Wiech, K., Kalisch, R., Weiskopf, N., Pleger, B., Stephan, K. E., & Dolan, R. J. (2006). Anterolateral Prefrontal Cortex Mediates the Analgesic Effect of Expected and Perceived Control over Pain. *The Journal of Neuroscience*, *26*(44), 11501–11509. https://doi.org/10.1523/JNEUROSCI.2568-06.2006

Willis, W. D., Iggo, A., Iversen, L. L., & Cervero, F. (1985). Nociceptive pathways: Anatomy and physiology of nociceptive ascending pathways. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *308*(1136), 253–268. https://doi.org/10.1098/rstb.1985.0025

Witt, I., & Griffin, J. P. (1962). Afferent Cutaneous C-fibre Reactivity to Repeated Thermal Stimuli. *Nature*, *194*(4830), Art. 4830. https://doi.org/10.1038/194776a0

Xuejing, L. U., & Xin, H. O. U. (2019). Predictive coding in auditory cortex: The neural responses to sound repetition and auditory change. *Advances in Psychological Science*, *27*(12), 1996. https://doi.org/10.3724/SP.J.1042.2019.01996

Zeki, S., & Shipp, S. (1988). The functional logic of cortical connections. *Nature*, *335*, 311–317. https://doi.org/10.1038/335311a0

Zelano, C., Mohanty, A., & Gottfried, J. A. (2011). Olfactory Predictive Codes and Stimulus Templates in Piriform Cortex. *Neuron*, *72*(1), 178–187. https://doi.org/10.1016/j.neuron.2011.08.010

Zhao, C., Widmer, Y. F., Diegelmann, S., Petrovici, M. A., Sprecher, S. G., & Senn, W. (2021). Predictive olfactory learning in Drosophila. *Scientific Reports*, *11*(1), Art. 1. https://doi.org/10.1038/s41598-021-85841-y

## 9.0 Zusammenfassung

In dieser Dissertation haben wir die Rolle der Bayesschen Inferenz in der Schmerzwahrnehmung unter Verwendung von *Predictive Coding* und *Active Inference* als theoretischen Rahmen erkundet. Wir untersuchten, wie das Gehirn Erfahrungen und Sinnesinformationen nutzt, um Vorhersagen über Schmerzerlebnisse zu treffen und wie diese Informationen integriert werden, um ein Schmerzempfinden in einem *Bayesschen Schmerzmodell* zu bilden. Indem wir Schmerzen aus einer Bayesschen Perspektive betrachten, konnten wir unser Verständnis der zugrundeliegenden neuralen Mechanismen der Schmerzwahrnehmung vertiefen und zur Entwicklung neuer Schmerzbewältigungs-Strategien beitragen. Insgesamt zeigt diese Arbeit wichtige Elemente, die über die Schmerzforschung hinaus zum Verständnis des *Bayesschen Gehirns* beitragen.

Unsere empirischen Arbeiten zeigen, dass das Gehirn Schmerzen über *Predictive Coding*-Mechanismen verarbeitet, indem es rekurrente Übertragungen von top-down Erwartungen und resultierenden bottom-up Erwartungsfehlern nutzt, die in spezifischen spektralen und temporalen Mustern kodiert werden. Die Untersuchung einer affektiv-visuellen Modalität liefert Evidenz dafür, dass diese oszillatorischen Prozesse in modalitätsspezifischen oszillatorischen Mustern widergespiegelt werden. Unsere Daten stützen die Auffassung, dass ein Schmerzempfinden in einer Bayes-optimalen Weise erzeugt wird, wie durch ein *Bayessches Schmerzmodell* postuliert. In zwei experimentellen Studien wird dies demonstriert durch bidirektionale Modulationen des Schmerzempfindens durch Placebo- und Nocebo-Erwartungen. Das *Bayessche Schmerzmodell* kann kontextuelle Modulationen durch eine Verschiebung der Intensitätserwartungen erklären, etwa eine Schmerzreduktion durch eine Selbstbehandlung. Dies steht im Widerspruch zu Ideen von *Active Inference*, die besagen, dass es zu einer Abschwächung der sensorischen Präzision durch eine Selbstbehandlung von Schmerz kommen sollte. Diese Dissertation zeigt die Anwendung von Bayesschen Prinzipien in der Schmerzverarbeitung, indem das Gehirn als statistische Maschine dargestellt wird, die optimale Inferenzen über die Welt vornimmt.

## 10.0 Zusammenfassung (Englisch)

In this dissertation, we have explored the role of Bayesian inference in the perception of pain, using *Predictive Coding* and *Active Inference* as theoretical frameworks. We have examined how the brain uses experience and sensory information to make predictions about potential pain experiences, and how this information is integrated to form a pain percept in a *Bayesian Pain Model*. By considering pain from a Bayesian perspective, we aimed to deepen our understanding of the underlying neural mechanisms of pain perception and to contribute to the development of new pain management strategies. Overall, this work highlights important elements that contribute to the understanding of the *Bayesian Brain* in pain.

Our empirical work demonstrates that the brain processes pain via *Predictive Coding* mechanisms, by utilizing recurrent transmissions of top-down expectations and bottom-up prediction errors, encoded in specific temporal and spectral patterns. The investigation of an affective-visual modality provides evidence that these oscillatory processes are reflected in modality-specific oscillatory patterns. Our data support the notion that a pain percept is generated in a Bayes-optimal manner, as explained by a *Bayesian Pain Model*. This is demonstrated in two experimental studies where we performed bi-directional modulations of the pain percept by placebo and nocebo expectations. The *Bayesian Pain Model* can explain contextual modulations via a mean shift in intensity expectations, for example, during a reduction in pain by a feeling of agency. This is in contradiction to ideas of *Active Inference*, which posits that there should be an attenuation of sensory precision by agency. This dissertation demonstrates the application of Bayesian principles in pain processing, framing the brain as a statistical machine that performs optimal inferences about the world.

## 11.0 Abkürzungsverzeichnis / List of Abbrevations

| | |
|---|---|
| EEG | Electroencephalography |
| MEG | Magentoencephalography |
| VBA | Variational Bayesian Analysis |
| VAS | Visual Analogue Scale |
| TENS | Transcutaneous Electric Nerve Stimulation |
| fMRI | Functional Magnetic Resonance Imaging |
| IAPS | International Affective Picture System |
| ICA | Independent Component Analysis |
| PCA | Patient-Controlled Analgesia |

## 12.0 Publikationsliste

Strube, A., Rose, M., Fazeli, S., & Büchel, C. (2021). The Temporal and Spectral Characteristics of Expectations and Prediction Errors in Pain and Thermoception. eLife, 10, e62809.

Strube, A., Rose, M., Fazeli, S., & Büchel, C. (2021). Alpha-to-Beta and Gamma-Band Activity Reflect Predictive Coding in Affective Visual Processing. Scientific Reports, 11(1), 1-15.

Strube, A., Horing, B., Rose, M., & Büchel, C. (2022). Agency Affects Pain Inference through Intensity Shift as Opposed to Precision Modulation [Manuscript submitted for publication]. Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf

## 13.0 Lebenslauf

Persönliche Angaben:

Andreas Strube
Geburtsdatum: 12.05.1994
Geburtsort: Hamburg

Ausbildung:

| | |
|---|---|
| 06/2018 – 01/2023 (vsl.) | Promotionsstudium (Ph.D.),<br>Universitätsklinikum Hamburg-Eppendorf |
| 10/2015 – 03/2018 | M.Sc. Cognitive Science,<br>Universität Osnabrück |
| 10/2012 – 06/2016 | B.Sc. Psychologie,<br>Universität Hamburg |

Berufserfahrung:

| | |
|---|---|
| Seit 03/2018 | Wissenschaftlicher Mitarbeiter,<br>Universitätsklinikum Hamburg-Eppendorf |
| 09/2016 – 02/2018 | Studentische Hilfskraft,<br>Universitätsklinikum Hamburg-Eppendorf |
| 10/2014 – 09/2015 | Studentische Hilfskraft,<br>Universität Hamburg |

## 14.0 Danksagung

Mein Dank gilt allen, die mich in den letzten 5 Jahren bei dieser Arbeit und in meinem Leben unterstützt haben. Zuallererst möchte ich meinen engsten Freunden, die mich vor allem vor allem in den zuletzt schwierigen Zeiten bedingungslos begleitet haben, meine tiefste Dankbarkeit aussprechen. Für mich wart ihr wie eine Familie seit meiner Kindheit: wichtiger als jeder Mentor, die größte Motivation und der liebste Grund zu sein. Ebenso bedanke ich mich bei meiner Familie; besonders danke ich meinem Bruder Alexander Strube für die gemeinsamen Abende während der Corona Pandemie. Wir haben uns gegenseitig gebraucht, und es ist ein großes Glück, dass wir diese Zeit miteinander verbringen durften.

Im Zusammenhang mit der Arbeit möchte ich mich bei meinen Kollegen und vor allem meiner Arbeitsgruppe bedanken: für das gute Arbeitsklima, für die Kollegialität, für die Akzeptanz und für das Verstandenwerden. Ein großer Dank gilt dabei Björn Horing, für deine Hilfe und Unterstützung, jederzeit, und für das beste (und strengste) Lektorat, das ich kenne. Ein besonderer Dank geht an Prof. Dr. Michael Rose, der mir in den emotionalsten und kompliziertesten Momenten meiner Doktorarbeit ein Vertrauter und Berater gewesen ist. Und schließlich möchte ich Prof. Dr. Christian Büchel danken, durch den ich als studentische Hilfskraft in die Forschungsgruppe kam. Danke dafür, dass du mir den Beginn meiner Promotion ermöglicht hast und dass du mich in schwierigen Zeiten bedingungslos unterstützt hast. Ich bin sehr dankbar für die Möglichkeiten, die du mir gegeben hast, für die Freiheit, mich zu entwickeln, und für dein aufrichtiges Zuhören, wenn ich es brauchte.

Vielen Dank.

## 15.0 Reprints

## 15.1 Study 1: The Temporal and Spectral Characteristics of Expectations and Prediction Errors in Pain and Thermoception

Strube, A., Rose, M., Fazeli, S., & Büchel, C. (2021). The Temporal and Spectral Characteristics of Expectations and Prediction Errors in Pain and Thermoception. eLife, 10, e62809.

RESEARCH ADVANCE

# The temporal and spectral characteristics of expectations and prediction errors in pain and thermoception

**Andreas Strube*, Michael Rose, Sepideh Fazeli, Christian Büchel***

Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

**Abstract** In the context of a generative model, such as predictive coding, pain and heat perception can be construed as the integration of expectation and input with their difference denoted as a prediction error. In a previous neuroimaging study (Geuter et al., 2017) we observed an important role of the insula in such a model but could not establish its temporal aspects. Here, we employed electroencephalography to investigate neural representations of predictions and prediction errors in heat and pain processing. Our data show that alpha-to-beta activity was associated with stimulus intensity expectation, followed by a negative modulation of gamma band activity by absolute prediction errors. This is in contrast to prediction errors in visual and auditory perception, which are associated with increased gamma band activity, but is in agreement with observations in working memory and word matching, which show gamma band activity for correct, rather than violated, predictions.

## Introduction

It has been shown that physically identical nociceptive input can lead to variable sensations of pain, depending on contextual factors (***Tracey and Mantyh, 2007***). In particular, attention, reappraisal, and expectation are core mechanisms that influence how nociception leads to pain (***Wiech et al., 2008***). A clinically important example of how expectations can shape pain processing is placebo hypoalgesia: pain relief mediated by expectation and experience – in the absence of active treatment (***Petrovic et al., 2002***; ***Wager et al., 2004***; ***Colloca and Benedetti, 2005***; ***Bingel et al., 2006***; ***Atlas and Wager, 2012***; ***Anchisi and Zanon, 2015***).

In the context of a generative model of pain, it has been proposed that pain perception can be seen as the consequence of an integration of expectations with nociception (***Büchel et al., 2014***; ***Wiech, 2016***; ***Ongaro and Kaptchuk, 2019***). In this framework, expectations are integrated with incoming nociceptive information and both are weighted by their relative precision (***Grahl et al., 2018***) to form a pain percept. This can be seen in analogy to ideas in multisensory integration (***Ernst and Banks, 2002***). Expectations or predictions and resulting prediction errors also play a key role in generative models such as predictive coding (***Huang et al., 2011***). In essence, this framework assumes that neuronal assemblies implement perception and learning by constantly matching incoming sensory data with the top-down predictions of an internal or generative model (***Knill and Pouget, 2004***; ***Huang et al., 2011***; ***Clark, 2013***). Basically, minimizing prediction errors allows systems to resist their tendency to disorder by the creation of models with better predictions regarding the sensory environment, leading to a more efficient encoding of information (***Friston, 2010***).

Electroencephalogram (EEG) correlates of nociceptive skin stimulation have been widely investigated. Generally, phasic gamma activity has been associated with stimulus intensity over the sensory cortex where the amplitudes of pain-induced gamma oscillations increase with objective stimulus intensity and subjective pain intensity (***Gross et al., 2007***; ***Hauck et al., 2007***; ***Zhang et al., 2012***;

***For correspondence:**
a.strube@uke.de (AS);
buechel@uke.de (CB)

**Competing interests:** The authors declare that no competing interests exist.

**Funding:** See page 17

**Received:** 10 September 2020
**Accepted:** 16 February 2021
**Published:** 17 February 2021

**Reviewing editor:** Peter Kok, University College London, United Kingdom

*Rossiter et al., 2013*; *Tiemann et al., 2015*). Additionally, pain-related gamma band oscillations have been linked to the insular cortex as well as temporal and frontal regions using depth electrodes in epilepsy patients (*Liberati et al., 2018*). In tonic painful heat stimulation, medial prefrontal gamma activity has been observed (*Schulz et al., 2015*). In addition, gamma activity is enhanced by attention in human EEG experiments in visual (*Gruber et al., 1999*), auditory (*Tiitinen et al., 1993*; *Debener et al., 2003*), and sensorimotor processing (i.e. tactile stimuli) (*Bauer et al., 2006*) as well as in nociception (*Hauck et al., 2007*; *Hauck et al., 2015*; *Tiemann et al., 2010*).

Pain-related alpha-to-beta band oscillations are typically found to be suppressed with higher stimulus intensity (*Mouraux et al., 2003*; *Ploner et al., 2006*; *May et al., 2012*; *Hu et al., 2013*), which is enhanced by attention (*May et al., 2012*) and (placebo) expectation (*Huneke et al., 2013*; *Tiemann et al., 2015*; *Albu and Meagher, 2016*). Interestingly, prestimulus theta (*Taesler and Rose, 2016*) as well as prestimulus alpha and gamma activity (*Tu et al., 2016*) can affect subsequent pain processing. Specifically, trials with smaller prestimulus alpha and gamma oscillations were perceived as more painful, suggesting a negative modulation of subsequent pain perception (*Tu et al., 2016*).

Cued pain paradigms (*Atlas et al., 2010*) have been used to generate expectations and prediction errors. Previous functional magnetic resonance imaging (fMRI) studies have suggested an important role of the anterior insular cortex for mediating expectation effects and the integration of prior expectation and prediction errors in the context of pain (*Ploghaus et al., 1999*; *Koyama et al., 2005*; *Atlas et al., 2010*; *Geuter et al., 2017*; *Fazeli and Büchel, 2018*). These studies have revealed that neuronal signals in the anterior insula represent predictions and prediction errors with respect to pain, which in theory would allow the combination of both terms as required for predictive coding (*Büchel et al., 2014*; *Ongaro and Kaptchuk, 2019*). However, in fMRI studies, predictions and prediction errors cannot be temporally dissociated due to the low temporal resolution of the method. To investigate this further, we conducted a cue-based pain experiment using EEG to achieve high temporal and spectral resolution of predictions and prediction error processes in the context of pain.

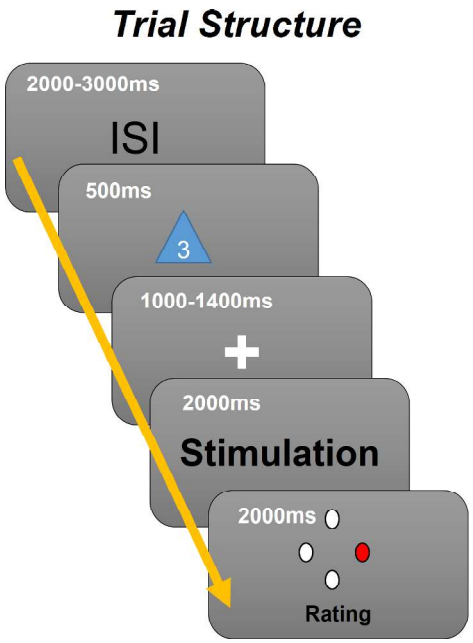
In this experiment (N = 29) we employed contact heat stimuli with three different intensities (low heat, medium heat, and high heat), preceded by a visual cue indicating the upcoming intensity (*Figure 1*). To generate prediction errors, the modality (picture or heat) was correctly cued only in 70% of all trials, and stimulus intensities were correctly cued only in 60% of all trials. We then investigated oscillatory activity related to stimulus intensity, expectation, and prediction errors (*Figure 2*).

Based on the previous data, we hypothesized that expectation signals should temporally precede prediction error signals. Based on the functional neuroanatomy of cortical microcircuits (*Bastos et al., 2012*), with feedforward connections predominately originating from superficial layers and feedback connections from deep layers, we expect that prediction error signals should be related to higher frequencies (e.g. gamma band) than prediction signals (*Todorovic et al., 2011*; *Arnal and Giraud, 2012*).

# Materials and methods

## Participants

We investigated 35 healthy male participants (mean 26, range 18–37 years), who were paid as compensation for their participation. Applicants were excluded if one of the following exclusion criteria applied: neurological, psychiatric, dermatological diseases, pain conditions, current medication, or substance abuse. All volunteers gave their informed consent. The study was approved by the Ethics Board of the Hamburg Medical Association. Of 35 participants, data from six participants had to be excluded from the final EEG data analysis due to technical issues during the EEG recording (i.e. the data of the excluded participants were contaminated with excessive muscle and/or technical artifacts) leaving a final sample of 29 participants. The sample size was determined according to a power calculation (G*Power V 3.1.9.4) based on *Geuter et al., 2017*. For the left anterior insula (fMRI; Table 1 in *Geuter et al., 2017*), we observed an effect size of partial eta squared of 0.17 and an effect size of 0.22 for the right anterior insula (cue × stimulus interaction). Using a power of (1-




**Cue-Stimulus Contingencies**

| | 42°C | 46°C | 48°C | Picture Level 1 | Picture Level 2 | Picture Level 3 |
|---|---|---|---|---|---|---|
| 1 | 7.0% | 2.3% | 2.3% | 3.0% | 1.0% | 1.0% |
| 2 | 2.3% | 7.0% | 2.3% | 1.0% | 3.0% | 1.0% |
| 3 | 2.3% | 2.3% | 7.0% | 1.0% | 1.0% | 3.0% |
| 1 | 3.0% | 1.0% | 1.0% | 7.0% | 2.3% | 2.3% |
| 2 | 1.0% | 3.0% | 1.0% | 2.3% | 7.0% | 2.3% |
| 3 | 1.0% | 1.0% | 3.0% | 2.3% | 2.3% | 7.0% |

**Figure 1.** Left: Graphical representation of the trial structure. Each trial started with the presentation of a cue, indicating the stimulus intensity and modality of the following stimulus. After a jittered phase where only the fixation cross was shown, the stimulus (visual or thermal) was presented. A rating phase (1–4) of the stimulus aversiveness followed. Right: Contingency table for all conditions for each cue–stimulus combination. Note that percentages are for all trials; therefore, each row adds up to 1/6 (six different cues).
The online version of this article includes the following figure supplement(s) for figure 1:

**Figure supplement 1.** Histogram showing the distribution of the total number of rejected components based on detected muscle artifacts.

beta) of 0.95 and an alpha level of 0.05 and assuming a low correlation (0.1) between repeated measures, this leads to a sample size of 25, assuming the weaker effect in the left insula.

## Stimuli and task

Stimulus properties were chosen to be identical to a previous fMRI study of predictive coding in pain where both expectation and absolute prediction error effects were observed (***Fazeli and Büchel, 2018***). Thermal stimulation was performed using a 30 × 30 $mm^2$ Peltier thermode (CHEPS Pathway, Medoc) at three different intensities: low heat (42˚C), medium heat (46˚C), and high heat (48˚C) at the left radial forearm. These three temperatures cover a large range of temperatures associated with nociception. The lowest temperature was set at 42˚C to ensure a temperature above the median threshold of heat-sensitive C-fiber nociceptors which have a median heat threshold of 41˚C (***Treede et al., 1998***). The baseline temperature was set at 33˚C and the rise rate to 40˚C/s. After two blocks, the stimulated skin patch was changed to avoid sensitization.

Aversive pictures were chosen from the International Affective Picture System (IAPS) (***Lang et al., 2008***) database at three different levels of aversiveness. The images presented during the EEG experiment had three levels of valence of which the low valence category had valence values of 2.02 ± 0.05 (mean ± standard error), the medium valence category had valence values of 4.06 ± 0.02 (mean ± standard error), and the high valence category had valence values of 5.23 ± 0.01 (mean ± standard error).

Prior to each picture or heat stimulus, a visual cue was presented. The color of the cue (triangle) indicated (probabilistically) the modality of the stimulus (orange for picture and blue for heat). A white digit written inside of each triangle indicated (probabilistically) the intensity of the subsequent stimulus (1, 2, and 3 for low, medium, and high intensity, respectively). During the whole trial, a centered fixation cross was presented on the screen.

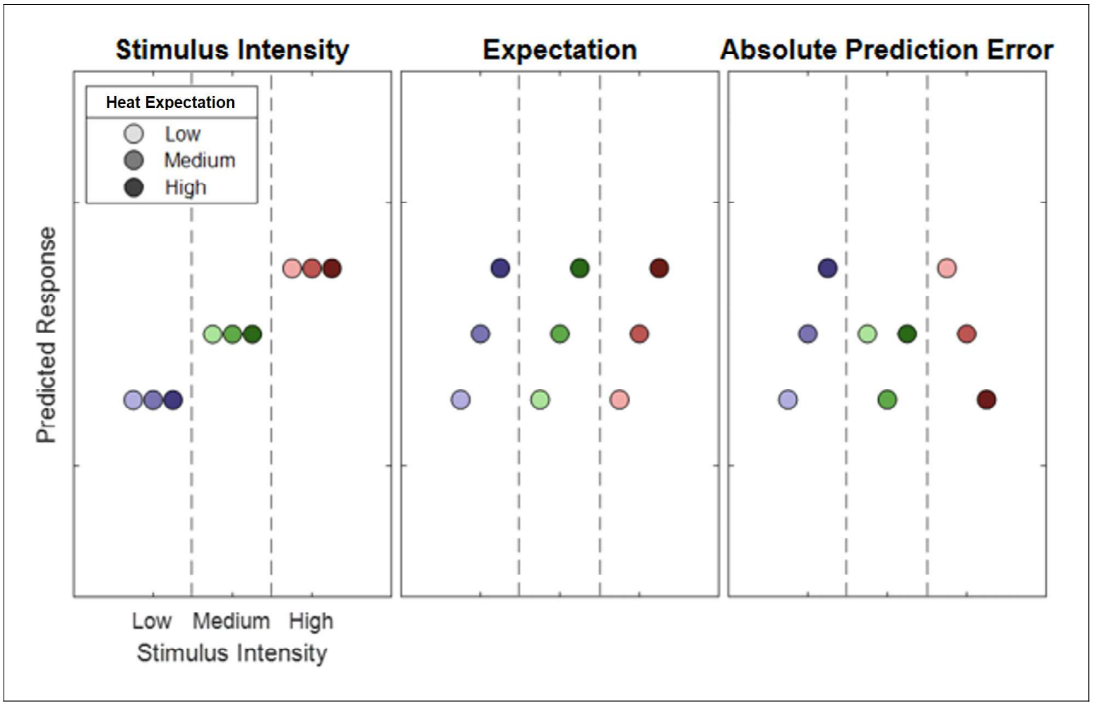



**Figure 2.** Hypothetical response patterns based on stimulus intensity (left), expectation (middle), and absolute prediction error (right). The y-axis represents a hypothetical response variable (e.g. electroencephalogram [EEG] power). Each dot represents a different condition for each stimulus–cue combination. Blue colors represent low heat conditions, green colors represent medium heat conditions, and red colors represent high heat conditions. Color intensities depict expectation level.

Each trial began with the presentation of the cue for 500 ms as an indicator for the modality and intensity of the subsequently presented stimulus. The modality was correctly cued in 70% of all trials by the color of the triangle. In 60% of all trials, the stimulus intensity was correctly indicated by the digit within the triangle (see ***Figure 1*** for an overview of all cue contingencies).

Before the presentation of the stimulus, there was a blank period with a variable time frame between 1000 and 1400 ms. Then, the visual or thermal stimulus was presented for a duration of 2 s. The visual stimulus was centered on the screen and allowed the participant to perceive the stimulus without eye movements. Right after the termination of the stimulus, subjects were asked to rate the aversiveness of the stimulus on a four-point rating scale, where one was labeled as 'neutral' and four was labeled as 'very strong'. Ratings were performed using a response box operated with the right hand (see ***Figure 1*** for a visualization of the trial structure).

In addition, four catch trials were included in each block. Subjects were asked to report the preceding cue in terms of their information content of the modality and intensity within 8 s, and no stimulation was given in these trials.

Trials were presented in four blocks. Each block consisted of 126 trials and four catch trials and lasted about 15 min. The trial order within each block was pseudorandomized. The order of blocks was randomized across subjects. The whole EEG experiment including preparation and instructions lasted for about 3 hr.

Prior to the actual EEG experiment, subjects participated in a behavioral training session. During this session, participants were informed about the procedure and gave their written informed consent. The behavioral training session was implemented to avoid learning effects associated with the contingencies between the cues and the stimuli during the EEG session. Between two and three blocks were presented during the training session (without electrophysiological recordings). The experimenter assessed the performance after each block based on the percentage of successful catch trials and the ability to distinguish the three levels of aversiveness of each modality. The training session was terminated after the second block if participants were able to successfully label cues in 75% of the catch trials within the second block.

## EEG data acquisition

EEG data were acquired using a 64-channel Ag/AgCl active electrode system (ActiCap64; BrainProducts) placed according to the extended 10–20 system (***Klem et al., 1999***). Sixty electrodes were used of the most central scalp positions. The EEG was sampled at 500 Hz, referenced at FCz, and grounded at Iz. For artifact removal, a horizontal, bipolar electrooculogram (EOG) was recorded using two of the remaining electrodes and placing them on the skin approximately 1 cm left from the left eye and right from the right eye at the height of the pupils. One vertical EOG was recorded using one of the remaining electrodes centrally approx. 1 cm beneath the left eyelid and another electrode was fixated on the neck at the upper part of the left trapezius muscle (Musculus trapezius) to record an electromyogram.

## EEG preprocessing

The data analysis was performed using the Fieldtrip toolbox for EEG/MEG (magnetoencephalogram) analysis (***Oostenveld et al., 2011***) at Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, the Netherlands (see http://www.ru.nl/neuroimaging/fieldtrip). EEG data were epoched and time-locked to the stimulus onset using the electrical trigger signaling the thermode to start the temperature rise of a given heat trial. Each epoch was centered (subtraction of the temporal mean) and detrended and included a time range of 3410 ms before and 2505 ms after trigger onset.

The data was band pass-filtered at 1–100 Hz, Butterworth, fourth order. EEG epochs were then visually inspected and trials contaminated by artifacts due to gross movements or technical artifacts were removed. Subsequently, trials contaminated by eye-blinks and movements were corrected using an independent component analysis (ICA) algorithm (***Makeig et al., 1996***; ***Jung et al., 2000***). In all datasets, individual eye movements, showing a large EOG channel contribution and a frontal scalp distribution, were clearly seen in the removed independent components. Additionally, time–frequency decomposed ICA data were inspected at a single trial level, after z-transformation (only for artifact detection purposes) based on the mean and the standard deviation across all components separately for each frequency from 31 to 100 Hz. Time–frequency representations were calculated using a sliding window multi-taper analysis with a window of 200 ms length, which was shifted over the data with a step size of 20 ms with a spectral smoothing of 15 Hz. Artifact components or trials were easily visible and were compared with the raw ICA components. Specifically, single and separate muscle spikes were identified as columns or 'clouds' in time–frequency plots. Using this procedure, up to 30 components were removed before remaining non-artifactual components were back-projected and resulted in corrected data. Subsequently, the data was re-referenced to a common average of all EEG channels and the previous reference channel FCz was reused as a data channel (see ***Figure 1—figure supplement 1*** for a summary of rejected components per participant).

Before time–frequency transformations for data analysis were performed on the cleaned dataset, the time axis of single trials was shifted to create cue-locked and stimulus-locked data. Cue-locked data uses the onset of the cue as t = 0. Stimulus-locked data takes the ramp-up period of the thermode into account and sets t = 0 to the time point when the thermode reached the target temperature (225, 325, and 375 ms after trigger onset for low, medium, and high heat conditions, respectively). Frequencies up to 30 Hz (1–30 Hz in 1 Hz steps) were analyzed using a sliding Hanning-window Fourier transformation with a window length of 300 ms and a step size of 50 ms. For the analysis of frequencies higher than 30 Hz (31–100 Hz in 1 Hz steps), spectral analyses of the EEG data were performed using a sliding window multi-taper analysis. A window of 200 ms length was shifted over the data with a step size of 50 ms with a spectral smoothing of 15 Hz. Spectral estimates were averaged for each subject over trials. Afterward, a z-baseline normalization was performed based on a 500 ms baseline before cue onset. For cue-locked data, a time frame ranging from −650 to −150 ms was chosen as a baseline. A distance from the cue onset to the baseline period of 150 ms was set because of the half-taper window length of 150 ms, that is, data points between −150 and 0 ms are contaminated by the onset of the cue. For stimulus-locked trials, a variable cue duration (1500–1900 ms) and a variable stimulus offset based on the ramp-up time (225–375 ms) were additionally taken into account, resulting in an according baseline from −2950 to −2450 ms from stimulus onset. For the baseline correction of time–frequency data, the mean and standard deviation were estimated for the baseline period (for each subject–channel–frequency combination,

separately). The mean spectral estimate of the baseline was then subtracted from each data point, and the resulting baseline-centered values were divided by the baseline standard deviation (classical baseline normalization – additive model; see ***Grandchamp and Delorme, 2011***).

## Predictive coding model

Similar to a previous fMRI study (***Fazeli and Büchel, 2018***), our full model included three experimental within-subject factors (see ***Figure 2***). The stimulus intensity factor (INT) models the measured response with a simple linear function of the stimulus intensity (−1, 0, and 1 for low, medium, and high intensities, respectively). The expectation factor (EXP) was defined (see ***Figure 2***; center column) linearly from the intensity predicted by the cue. Again, conditions with a low intensity cue were coded with a −1, conditions with a medium intensity cue with a 0 and conditions with a high intensity cue with a 1. The absolute prediction error factor (PE) resulted from the absolute difference of the expectation and actual stimulus intensity (see ***Figure 2***; right column).

We also investigated a signed PE. However, it should be noted that such a term is not orthogonal to the EXP. However, assuming that an EXP can only be observed after the cue and a PE after the nociceptive stimulus, we were able to test for a signed PE during the heat phase. Also, we considered a one-sided PE, where a prediction error is only signaled when the stimulus is more intense as expected, which is motivated by previous work (***Egner et al., 2010***; ***Summerfield and de Lange, 2014***; ***Geuter et al., 2017***).

As the lowest stimulus intensity was often perceived as non-painful, we additionally performed an analysis only with medium and high stimulus intensities. Accordingly, the lowest stimulus intensity (42˚C) were excluded in an additional repeated-measures ANOVA for this purpose (which will be referred to as the *reduced pain model*).

## Behavioral aversiveness ratings

Behavioral aversiveness ratings were averaged for all 3 × 3 cue–stimulus combinations over each participant, resulting in a 29 × 9 matrix (subject × condition) for the full model and a 29 × 6 matrix for the reduced pain model. We tested for main effects across stimulus intensity, expectation, as well as prediction error using a repeated-measures ANOVA as implemented in MATLAB (see fitrm and ranova; version 2020a, The MathWorks).

## EEG data analysis

All statistical tests in electrode space were corrected for multiple comparisons using non-parametrical permutation tests of clusters (***Maris and Oostenveld, 2007***). Cluster permutation tests take into account that biological processes are not strictly locked to a single frequency or time point and that activity could be picked up by multiple electrodes. Cluster permutation tests are specifically useful for explorative testing, as explained by ***Maris and Oostenveld, 2007***. While prior hypotheses could have been formulated regarding the spatial, temporal, and spectral characteristics of brain responses associated with the intensity of thermal stimulation, and regions of interest could have been described, variations in the present design could be related to temporal and spectral differences compared to other studies, which would be taken into account using non-parametric cluster permutation testing.

We wanted to explore positive and negative time–frequency patterns associated with our variations of stimulus intensity, expectation, and absolute prediction errors using a repeated-measures ANOVA. A statistical value corresponding to a p-value of 0.05 (F[1,28] = 4.196) obtained from the repeated-measures ANOVA F-statistics of the respective main effect was used for clustering. Samples (exceeding the threshold of F[1,28] = 4.196) were clustered in connected sets on the basis of temporal (i.e. adjacent time points), spatial (i.e. neighboring electrodes), and spectral (i.e. +/− 1 Hz) adjacency. Further, clustering was restricted in a way that only samples were included in a cluster which had at least one significant neighbor in electrode space, that is, at least one neighboring channel also had to exceed the threshold for a sample to be included in the cluster. Neighbors were defined by a template provided by the Fieldtrip toolbox corresponding to the used EEG montage.

Subsequently, a cluster value was defined as the sum of all statistical values of included samples. Monte Carlo sampling was used to generate 1000 random permutations of the design matrix, and statistical tests were repeated in time–frequency space with the random design matrix. The

probability of a cluster from the original design matrix (p-value) was calculated by the proportion of random design matrices producing a cluster with a cluster value exceeding the original cluster. This test was applied two-sided for negative and positive clusters, which were differentiated by the average slope of the estimated factors.

Monte Carlo cluster tests were performed with 1000 permutations using the test statistics of a repeated-measures ANOVA model as the value for clustering (at $p<0.05/F[1,28]=4.196$). All tests were performed for low frequencies (1–30 Hz) and high frequencies (31–100 Hz), separately. Muscular and ocular electrodes were excluded from the cluster analysis.

The within-subject INT (which was coded as increasing linearly with stimulus intensity) was tested stimulus-locked from 0 to 1.6 s. The within-subject EXP, which was coded as increasing linearly with the cued stimulus intensity, was tested cue-locked from 0 to 3.6 s. The signed PE was coded as the difference between stimulus intensity and expectation. The absolute prediction error was coded as the absolute difference between stimulus intensity and expectation (see ***Figure 2*** for details). Additionally, we tested a one-sided prediction error, occurring only when the actual stimulus is of a higher intensity than expected. The signed, absolute, and one-sided PEs were tested stimulus-locked from 0 to 1.6 s.

# Results

## Behavioral data – aversiveness ratings

Participants experienced aversive heat or saw picture stimuli which were probabilistically cued in terms of modality and intensity, evoking an expectation of modality and intensity. The subsequently applied stimuli were then rated on a visual analog scale (VAS) from 1 to 4. Our primary behavioral question was whether ratings are influenced by the experimental manipulation of stimulus intensity, expectation, and absolute prediction errors.

To validate our intensity manipulation for thermal stimuli and to verify the discriminability between different levels of heat, we first tested for the main effect of stimulus intensity (***Figure 3a***). Our data show a clear rating difference between the three levels of heat. Results regarding the aversive pictures are not the focus of this report but are depicted in ***Figure 3b*** for the sake of comparison.

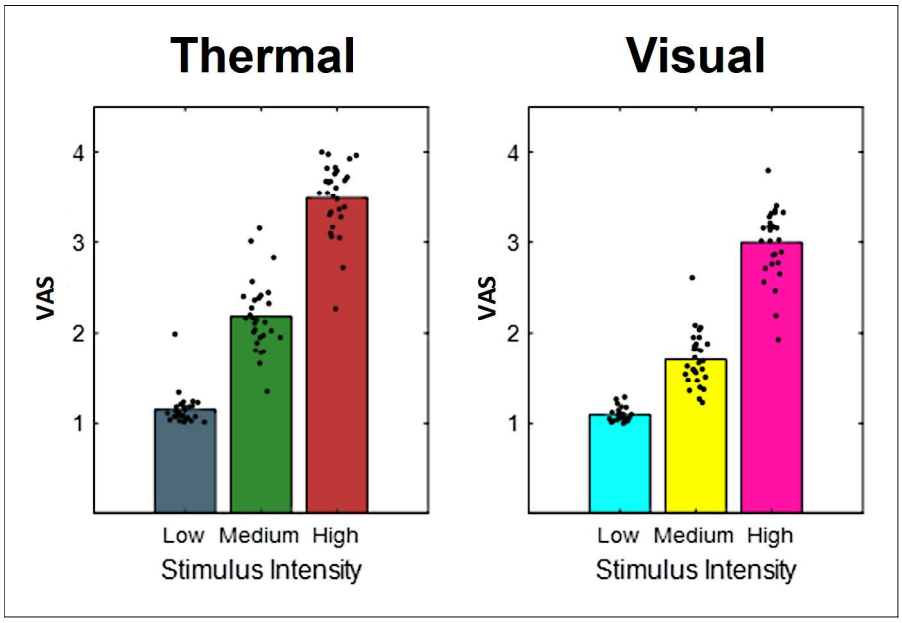



**Figure 3.** Bars indicate pooled aversiveness ratings for (**a**) heat and (**b**) aversive pictures for low-, medium-, and high-intensity conditions. Dots indicate average single-subject ratings.

To evaluate the main effects of stimulus intensity, expectation, and absolute prediction errors, we employed a repeated-measures ANOVA of the behavioral data, which revealed significant effects for the main effect of stimulus intensity, that is, the three levels of heat ($F[1,28] = 743.97$, $p<0.001$). Also, the main effect for expectation on aversiveness ratings was significant ($F[1,28] = 38.53$, $p<0.001$) (***Table 1***), indicating an influence of the cued intensity on behavioral aversiveness ratings (***Figure 4***). The absolute difference between the cued intensity and the actual stimulus intensity (i.e. absolute prediction error) only showed a trend effect on aversiveness ratings ($F[1,28] = 2.87$, $p=0.10$). The results regarding the aversive pictures are summarized in ***Table 1***.

## EEG – stimulus intensity

In a first EEG analysis, we tested for a main effect of the intensity of the heat input in the context of a correctly cued modality (i.e. heat was expected and received). In order to do so, we performed a repeated-measures ANOVA on the time–frequency representation of the EEG data on low frequencies (1–30 Hz) and high frequencies (31–100 Hz) separately after stimulus onset using a cluster correction criterion to address the multiple comparisons problem (see 'Materials and methods' for details). Any significant cluster – composed of neighboring data points in time, frequency, and space – would indicate a neuronal oscillatory representation of variations in stimulus intensity in a given frequency band.

In the low frequency range (1–30 Hz), one positive cluster (i.e. a positive average slope of the factor) and one negative cluster (i.e. a negative average slope of the factor) were significant (***Figure 5***), indicating a linear association of stimulus intensity and power in this frequency range. Specifically, the negative cluster included samples in a time range from 250 to 1600 ms after stimulus onset in a frequency range from 1 to 30 Hz, predominately at contralateral central electrode sites ($p=0.002$). The highest parametric F-value within this cluster obtained from the repeated-measures ANOVA was $F(1,28) = 36.40$ ($p<0.001$). This sample was observed at 1250 ms and 22 Hz and had a maximum at channel CP2. All channels included samples of the negative low frequency stimulus intensity cluster.

Also in the low frequency range (1–30 Hz), a positive significant cluster included samples in a time range from 150 to 1050 ms after stimulus onset in the theta frequency range from 1 to 7 Hz predominately at midline electrode sites ($p=0.048$). The highest parametric F-value from the repeated-measures ANOVA was $F(1,28) = 27.93$ ($p<0.001$). This sample was found at 550 ms and 3 Hz and had a maximum at channel O2. All channels except FC5, CP4, C6, and FT7 were part of the positive low frequency stimulus intensity cluster.

In the high frequency range (31–100 Hz) representing gamma activity, one positive cluster was significant ($p<0.001$). This cluster included samples in a time range from 550 to 1600 ms after stimulus onset and frequencies from 46 to 100 Hz, predominately at contralateral centroparietal electrode sites (***Figure 5***). The highest parametric F-value within this cluster obtained from the repeated-measures ANOVA was $F(1,28) = 33.35$ ($p<0.001$). This sample was observed at 1600 ms and 100 Hz and had a maximum at channel Cz. All channels included samples of the positive high frequency stimulus intensity cluster.

In conclusion, these results indicate that a higher intensity of the thermal input is associated with increased theta and gamma band power and a negative relationship of alpha-to-beta band power and the intensity of the thermal input (see ***Figure 5*** for a summary of the results of the main effect

**Table 1.** Main effects of stimulus intensity, expectation, and absolute prediction errors on subjective ratings in both heat and picture conditions.

| | Stimulus intensity (INT) | | Cued intensity (EXP) | | Absolute prediction error (PE) | |
|---|---|---|---|---|---|---|
| Factor | F(1,28) | p | F(1,28) | p | F(1,28) | p |
| Modality | | | | | | |
| Thermal | 743.97 | <0.001 | 39.53 | <0.001 | 2.87 | 0.10 |
| Visual | 762.10 | <0.001 | 1.46 | 0.24 | 7.7 | 0.01 |

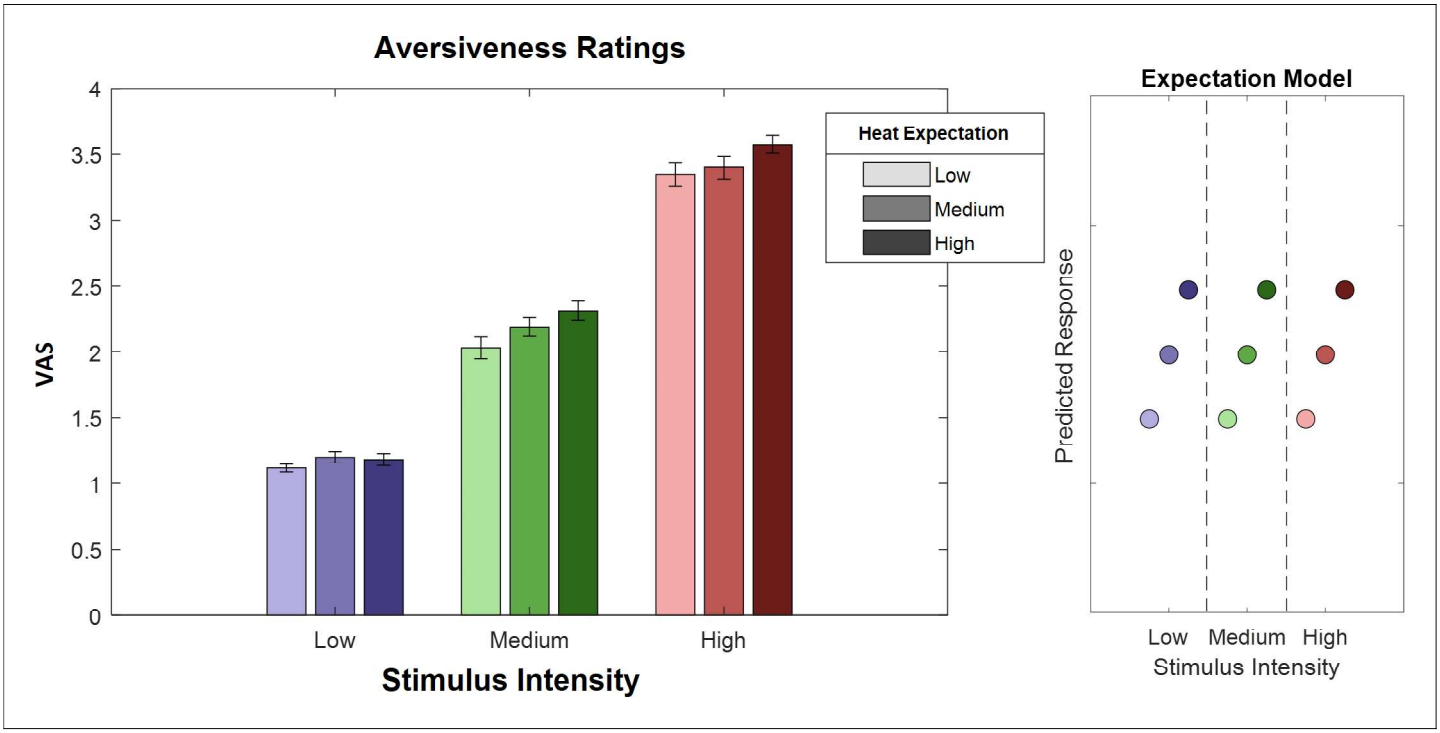



**Figure 4.** Ratings for heat stimuli (left) and 'expectation factor' weights (right). Bars indicate average aversiveness ratings. Ratings were given on a scale from 1 to 4. Error bars depict SEM. The data shows not only an effect of stimulus intensity (increase from blue to green to red) but also an effect of expectation (low to medium to high expectation). The right figure represents hypothetical response patterns based on the expectation factor. The y-axis represents the hypothetical response variable (e.g. visual analog scale [VAS] rating). Each dot represents a different condition for each stimulus–cue combination. Blue colors represent low heat conditions, green colors represent medium heat conditions, and red colors represent high heat conditions. Color intensities depict expectation level.

of stimulus intensity; see ***Figure 5—figure supplement 1*** for single-subject differences in the gamma band between low stimulus intensity and high stimulus intensity trials).

## Expectation

In a next step, we investigated the representation of EXP in our repeated-measures model, again for low frequencies (1–30 Hz) and high frequencies (31–100 Hz) separately in the cue-locked time–frequency representation of the EEG data.

This analysis revealed one significant positive cluster in the low frequency range (1–30 Hz), indicating a linear association of cue intensity (EXP) and power in this frequency range ($p<0.05$). The expectation cluster ($p=0.022$) included samples from time points ranging from 100 to 2000 ms after cue onset and included frequencies from 1 to 20 Hz. The highest parametric statistical test value ($F[1,28] = 26.96$, $p<0.001$) was observed at channel P1 700 ms after cue onset at a frequency of 9 Hz. All channels except TP8 included samples of the late expectation cluster (see ***Figure 6*** for a summary of the results of the expectation cluster; see ***Figure 6—figure supplement 1*** for single-subject values).

In summary, these results suggest an increase in alpha-to-beta band power to be associated with our experimental manipulation of expectations regarding the intensity of the thermal input.

## Prediction error model

Likewise, clustering was performed for the prediction error term after stimulus onset in low (1–30 Hz) and high frequencies (31–100 Hz). Any significant cluster would associate oscillatory activity with the difference of the expectation regarding the intensity of the thermal stimulation and the actual stimulation, representing a violation of this expectation (prediction error).

This analysis revealed a significant negative cluster in the high frequency range (31–100 Hz), indicating a (negative) linear association of absolute prediction errors and power in this frequency range

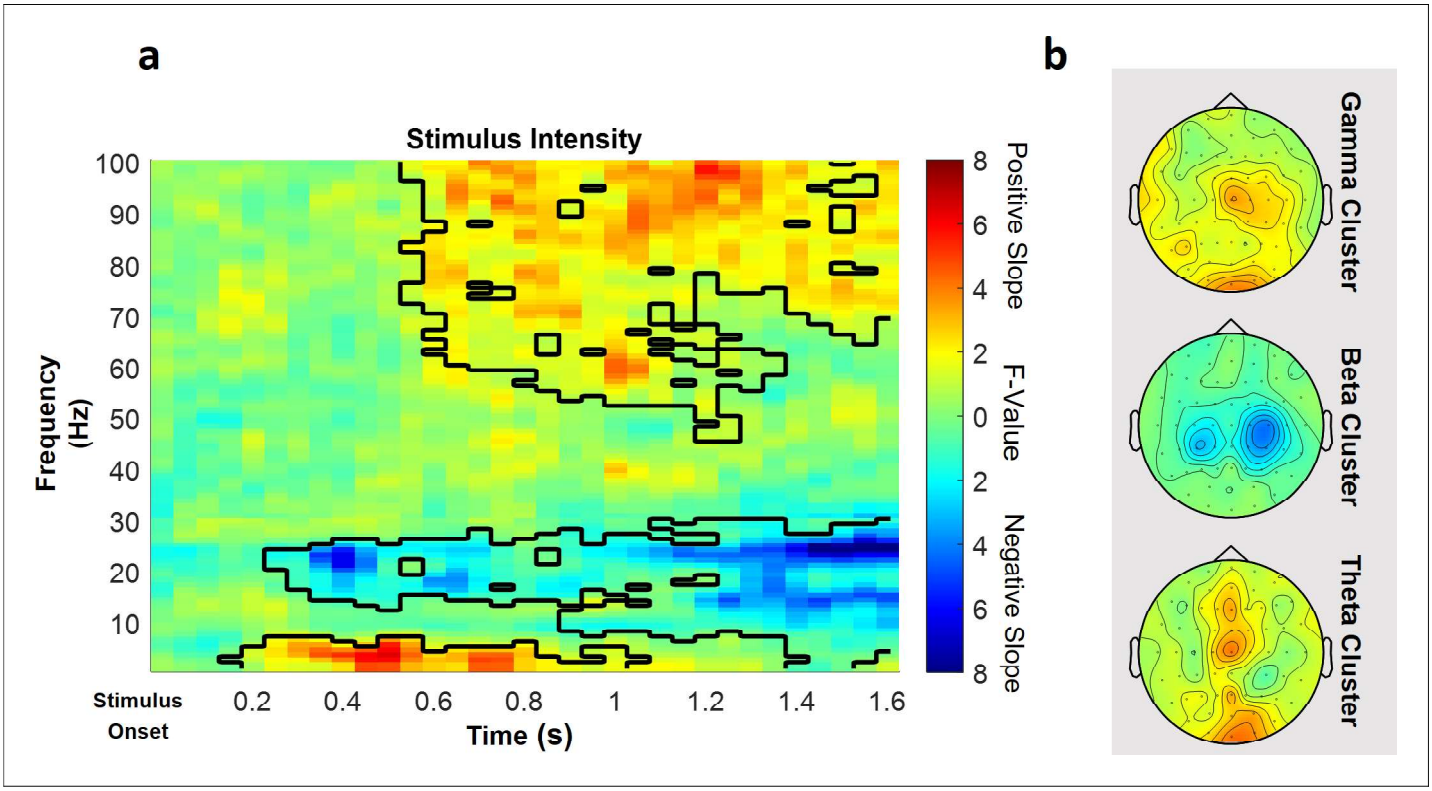



**Figure 5.** Parametric effects of stimulus intensity. Time–frequency representation averaged over all channels including a significant time–frequency sample of any cluster (**a**) and topographies over the whole cluster extents (i.e. full time and frequency range), respectively (**b**), of the stimulus intensity main effect of the repeated-measures ANOVA depicting increases (warm) and decreases (cold) in power in relation to heat stimulus intensity. Significant clusters are highlighted. Colors represent F-values from the repeated-measures ANOVA statistics for the main effect of stimulus intensity.
The online version of this article includes the following figure supplement(s) for figure 5:

**Figure supplement 1.** Difference for the main effect of stimulus intensity in the gamma band (averaged over 60–100 Hz, 1250–1600 ms) in power values for all high heat vs. low heat conditions with a valid modality cue (expect heat receive heat) for each subject, respectively.

(p=0.002). This (negative) absolute prediction error cluster included samples from frequencies ranging from 51 to 100 Hz and time points ranging from 50 to 1600 ms after stimulus onset. The highest parametric statistical test value ($F(1,28) = 28.52$, $p<0.001$) was found at channel O1 1300 ms after stimulus onset at a frequency of 98 Hz. All channels included samples of the absolute prediction error cluster (see ***Figure 7*** for a summary of the results; see ***Figure 7—figure supplement 1*** for single-subject values).

A cluster analysis of the signed prediction error, stimulus-locked after stimulus onset (from 1 to 30 Hz for low frequencies and 31–100 Hz for gamma frequencies; from 0 to 1600 ms, stimulus-locked), did not reveal any significant cluster of activity associated with a linear increase or decrease of EXP (all $p>0.05$). Ignoring all stimulus–cue combinations of the PE where the stimulus intensity was less intense than expected leads to a one-sided PE. A cluster analysis of this effect did not reveal any significant cluster of activity (all $p>0.05$).

In summary, these results suggest a decrease in gamma band power to be associated with our experimental manipulation of expectation violations, resulting from a mismatch of the cued intensity and the actual heat input.

## Reduced pain model

In an additional analysis, we tested all effects in a reduced pain model, which only included painful stimuli (i.e. three expectation levels and two intensity levels).

To evaluate the main effects of stimulus intensity, expectation, and absolute prediction errors in the behavioral data, we employed again a repeated-measures ANOVA which revealed significant

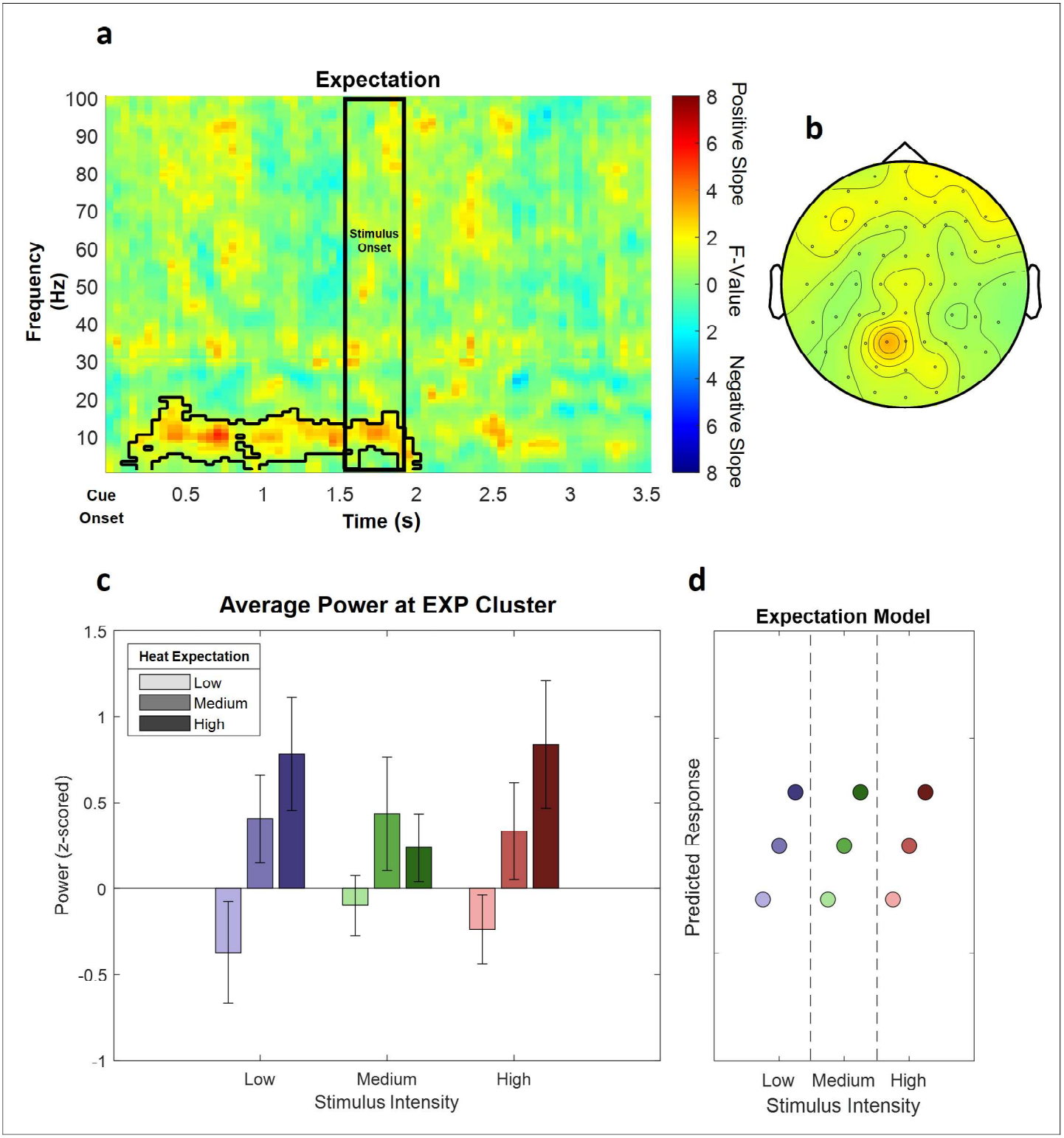



**Figure 6.** The main effect of expectation. (**a**) Time–frequency representation of the statistical F-values averaged over all channels. The significant cluster is highlighted. The black box between 1500 and 1900 ms marks the jittered onset of the trigger signal to start the ramp-up of the heat stimulus. (**b**) Topography of the averaged power over time and frequency of the whole cluster extent (i.e. over the whole time and frequency range) at each channel. Brighter colors indicate higher F-values. (**c**) Power values for all conditions with a valid modality cue (expect heat receive heat) averaged over all significant time–frequency–electrode samples of the EXP cluster show alpha-to-beta enhancement (i.e. positive representation) associated with expectation. Error bars represent SEM. (**d**) Predicted responses based on the positive expectation factor are shown. The y-axis represents an imaginary

*Figure 6 continued on next page*

*Figure 6 continued*

response variable (e.g. EEG power). Each dot represents a different condition (in the order of the bar plot representation of average EEG power) for each stimulus–cue combination. Blue colors represent low heat conditions, green colors represent medium heat conditions, and red colors represent high heat conditions. Color intensities depict expectation level.

The online version of this article includes the following figure supplement(s) for figure 6:

**Figure supplement 1.** Power values for all conditions with a valid modality cue (expect heat receive heat) averaged over all significant time–frequency–electrode samples period for each subject (ID) of the EXP cluster.

effects for the main effect of stimulus intensity, that is, the two remaining levels of pain ($F[1,28] = 1109.9$, $p<0.001$). Also, the main effect for expectation on pain ratings was significant ($F[1,28] = 17.07$, $p<0.001$), indicating again an influence of the cued intensity on behavioral pain ratings. The absolute difference between the cued intensity and the actual stimulus intensity (i.e. absolute prediction error) when only painful stimuli were included revealed a positive significant effect on pain ratings ($F[1,28] = 80.75$, $p<0.001$). This indicates prediction errors and prior expectations to modulate behavioral aversiveness ratings in painful stimulation.

For the analysis of time–frequency EEG data, we performed a repeated-measures ANOVA on the time–frequency representation of the EEG data on low frequencies (1–30 Hz) and high frequencies (31–100 Hz) separately and again using the same cluster correction criterion to address the multiple comparisons problem as in the initial analysis of the full model.

The cluster test of stimulus intensity revealed one negative cluster ($p=0.014$) in the low frequency range (1–30 Hz) including time points from 850 to 1600 ms and frequencies from 8 to 30 Hz (***Figure 8a***; see ***Figure 8—figure supplement 1*** for single-subject values). The maximum statistical F-value ($F[1,28] = 31.82$; $p<0.001$) was found at channel AF3 at a frequency of 30 Hz at 1600 ms and revealed a similar but more broad topography as compared to the original alpha-to-beta negative main effect of stimulus intensity of the analysis of the full model. All channels included samples of the negative stimulus intensity cluster.

In the high frequency range (31–100 Hz), a negative cluster of activity ($p=0.038$) was associated with absolute prediction errors and included samples in a time range from 850 to 1600 ms after stimulus onset in the gamma frequency range from 54 to 90 Hz predominately at occipital and parietal electrode sites. The highest parametric F-value from the repeated-measures ANOVA was $F(1,28) = 24.10$ ($p<0.001$). This sample was found at 1150 ms and 77 Hz and had a maximum at channel F8 (***Figure 8b***; see ***Figure 8—figure supplement 2*** for single-subject values). All channels except FC5, CP4, C6, and FT7 were part of the gamma frequency negative absolute prediction error cluster.

In the low (1–30 Hz) and high frequency (31–100 Hz) ranges, no significant cluster was observed representing a significant relationship between expectations and EEG activity. However, one cluster in the low frequency range (1–30 Hz) showed a trend level ($p=0.14$; based on cluster mass, i.e., the sum of all clustered F-values) and included samples in a time range from 550 to 1600 ms after stimulus onset and frequencies from 6 to 24 Hz and is displayed in dotted lines in ***Figure 8c*** (see ***Figure 8—figure supplement 3*** for single-subject values).

## Discussion

Using a cued heat paradigm with three different stimulus intensities, our data showed a clear discriminability of different levels of aversiveness based on behavioral ratings and EEG time–frequency patterns. Specifically, we observed several clusters of activity to be associated with the intensity of thermal stimulation in the theta, beta, and gamma band. Furthermore, behavioral data clearly indicated a positive influence of cued intensity on pain perception. In addition, our results provide evidence for temporally and spectrally separable clusters of oscillatory activity associated with expectation and a negative modulation of gamma activity by prediction errors for thermoception and pain. Specifically, one early low frequency (1–30 Hz) cluster was related to expectation in thermoception, that is, cued intensity. In contrast, a later occurring cluster at higher frequencies (31–100 Hz) was related to negative prediction errors in thermoception and pain.

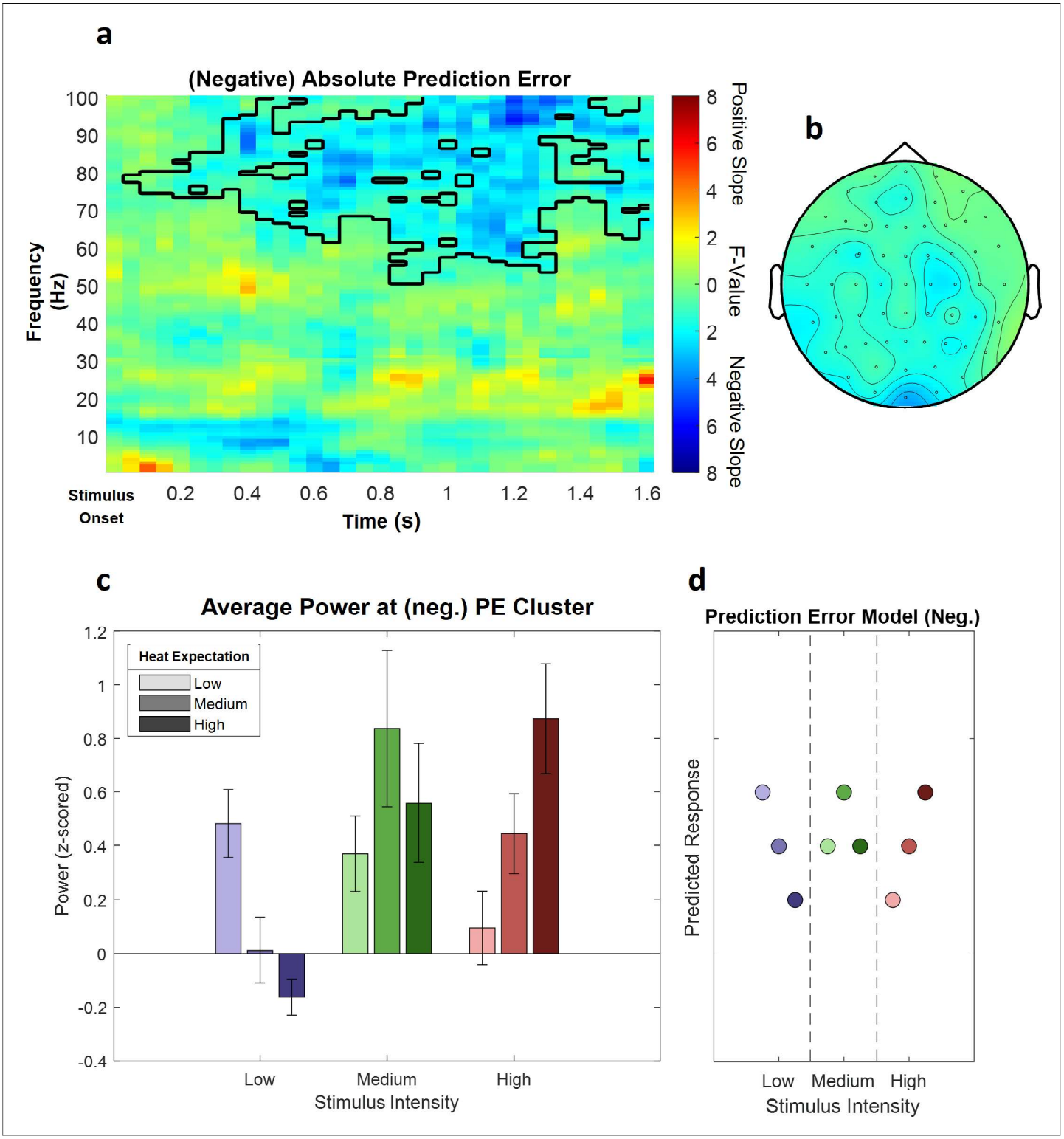



**Figure 7.** The main effect of absolute prediction errors. (**a**) Time–frequency representation of the statistical F-values averaged over all channels. The significant cluster is highlighted. (**b**) Topography of the averaged power over time and frequency of the whole cluster extent (i.e. over the whole time and frequency range) at each channel. Brighter colors indicate higher F-values. (**c**) Power values for all conditions with a valid modality cue (expect heat receive heat) averaged over all significant time–frequency–electrode samples of the prediction error factor (PE) cluster show gamma decreases (i.e. negative representation) associated with prediction errors. Error bars represent SEM. (**d**) Predicted responses based on the negative PE are shown: The y-axis represents an imaginary response variable (e.g. electroencephalogram [EEG] power). Each dot represents a different condition (in the order of

*Figure 7 continued on next page*

*Figure 7 continued*

the bar plot representation of average EEG power) for each stimulus–cue combination. Blue colors represent low heat conditions, green colors represent medium heat conditions, and red colors represent high heat conditions. Color intensities depict expectation level.
The online version of this article includes the following figure supplement(s) for figure 7:

**Figure supplement 1.** Power values for all conditions with a valid modality cue (expect heat receive heat) averaged over all significant time–frequency–electrode samples period for each subject (ID) of the negative absolute prediction error cluster.

## Stimulus intensity and oscillatory activity

Note that our definition of stimulus onset is based on the moment the thermode reached the target temperature. Using a thermode heating gradient of 40˚C/s and neglecting any small internal delays, the target temperatures of 42˚C, 46˚C, and 48˚C are reached after 225, 325, and 375 ms, respectively. Therefore, our observed increase in theta power agrees with previous studies (***Ploner et al., 2017***) and most likely correspond to pain-related evoked potentials (***Lorenz and Garcia-Larrea, 2003***; ***Tiemann et al., 2015***), such as the P2 with a similar topography. In addition, we observed a significant suppression of alpha-to-beta activity which, given the abovementioned delays of our painful stimuli, is in line with the reported beta suppression in previous EEG studies on pain (***Mouraux et al., 2003***; ***Ploner et al., 2006***; ***May et al., 2012***; ***Hu et al., 2013***). Finally, power in the gamma band was also correlated with heat intensity, which is in line with previous studies (***Gross et al., 2007***; ***Hauck et al., 2007***; ***Zhang et al., 2012***; ***Rossiter et al., 2013***; ***Tiemann et al., 2015***). Interestingly, only the alpha-to-beta band desynchronization differentiated between medium and high pain conditions, whereas differences in the theta and gamma band activity were only evident when the lowest stimulus intensity was included which was perceived as neutral.

We observed a behavioral effect of prediction errors on perceived stimulus intensity in the reduced pain model, but this effect was only a trend in the full model. The latter finding replicates a previous study (***Fazeli and Büchel, 2018***) indicating a robust effect. Interestingly, the effect of prediction errors on perception increased, and became significant, when we constrained our analysis to the clearly painful stimuli (reduced pain model). This suggests that a prediction error seems to more strongly affect pain perception, whereas the effect is weaker in the context of thermoception. However, this speculation should be corroborated in a future study.

On a more conceptual level, the investigation of neurophysiological effects even in the absence of a behavioral effect has been considered meaningful (***Wilkinson and Halligan, 2004***). In particular, the authors argue that because it is commonly unknown which parts of a cognitive process (and in which way) produce a specific behavioral response, the relationship between neurophysiological data and behavioral responses should not be overemphasized, and therefore it can be misleading to declare behavioral effects a reference or 'gold standard'. Studies aiming to understand neurophysiological mechanisms of cognition usually relate a neurophysiological readout to a known perturbation (i.e. experimental design), which is meaningful in its own right.

## Hypotheses based on microcircuits

Theoretical accounts (***Arnal and Giraud, 2012***; ***Bastos et al., 2012***) have suggested that predictive coding mechanisms could be related to the functional architecture of neuronal microcircuits. As feedforward connections are predominately originating from superficial layers and feedback connections from deep layers, it has been suggested that prediction errors should be expressed by higher frequencies than the predictions that accumulate them.

In the auditory modality, these ideas are supported by empirical data (***Todorovic et al., 2011***) showing that prediction errors in the context of repetition suppression were associated with higher gamma band activity. Likewise, in the visual domain, an MEG study has shown that temporo-parietal beta power was correlated with the predictability of an action kinematics–outcome sequence, while gamma power was correlated with the prediction error (***van Pelt et al., 2016***).

## Frequency patterns in predictive coding of pain

Only a few studies have investigated the spectral and temporal properties of expectations and prediction errors in the context of pain (summarized by ***Ploner et al., 2017***). A recent study in rodents has suggested an information flow between S1 gamma and ACC (Anterior Cingulate Cortex) beta

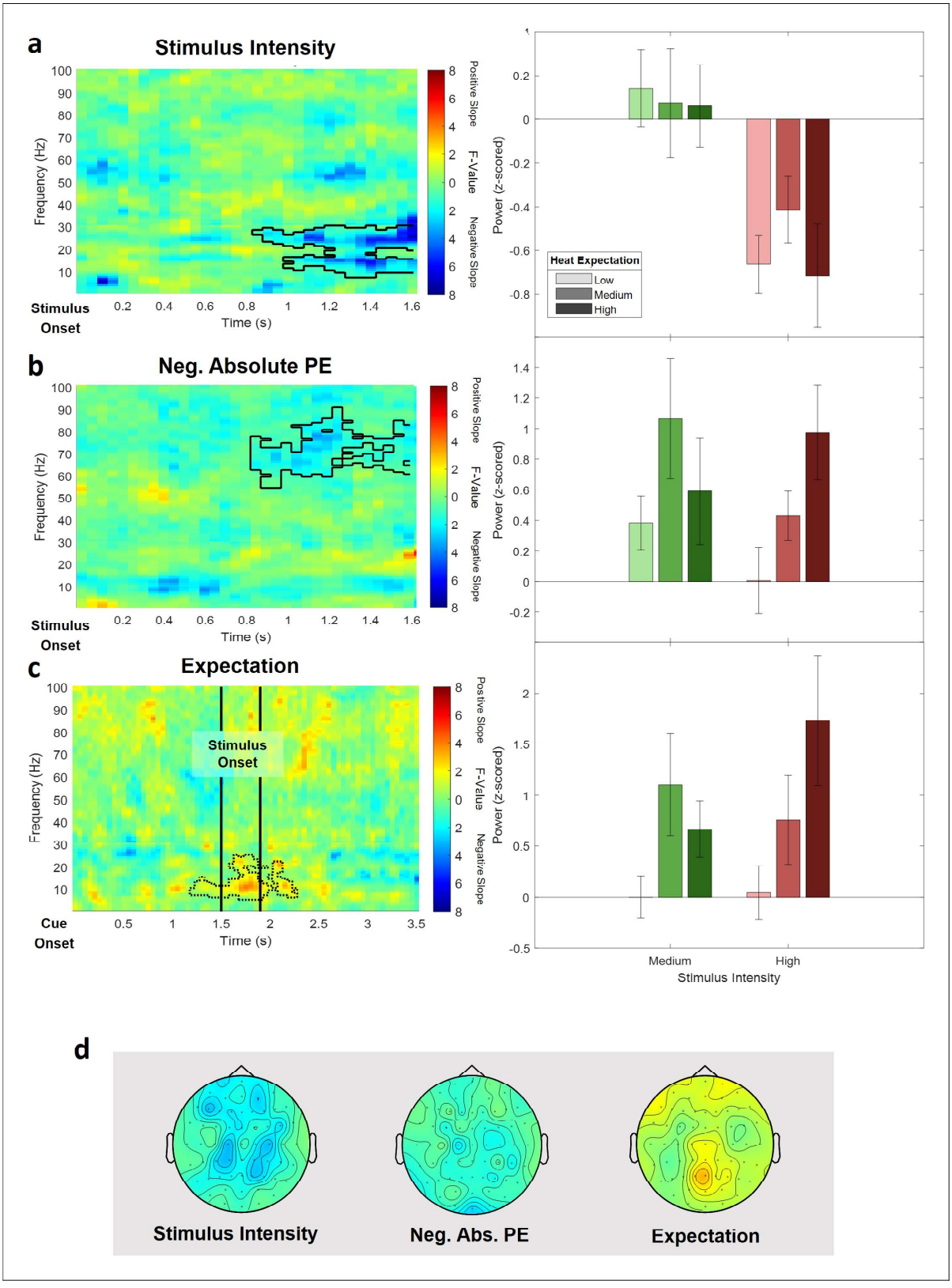



**Figure 8.** Electroencephalogram (EEG) data analysis of the reduced pain model. The top three rows show (**a**) the main effect of stimulus intensity, (**b**) the main effect of negative absolute prediction errors, and (**c**) the main effect of expectation. Left column: time–frequency representation of the statistical F-values averaged over all channels. Significant clusters are highlighted by a solid line. The non-significant expectation cluster is highlighted by a thin dotted line. Right column: power values for all conditions included in the reduced model with a valid modality cue (expect heat receive heat)

*Figure 8 continued on next page*

*Figure 8 continued*

averaged over all significant time–frequency–electrode samples of the respective cluster. (**d**) Topographies of the averaged power over time and frequency of the whole cluster extent (i.e. over the whole time and frequency range) at each channel for stimulus intensity (left), negative absolute prediction errors (center), and expectation (right). Brighter colors indicate higher F-values.
The online version of this article includes the following figure supplement(s) for figure 8:

**Figure supplement 1.** Power values for all medium and high intensity conditions with a valid modality cue.
**Figure supplement 2.** Power values for all medium and high intensity conditions with a valid modality cue.
**Figure supplement 3.** Power values for all medium and high intensity conditions with a valid modality cue.

activity during spontaneous pain (***Xiao et al., 2019***). Based on these data, the authors have proposed a predictive coding model including a bottom-up (gamma) and top-down (beta) component (***Song et al., 2019***). Finally, in humans, a recent EEG study showed that the sensorimotor cortex is more strongly connected to the medial prefrontal cortex at alpha frequencies during tonic pain, suggesting alpha band activity in tonic pain to be associated with bottom-up instead of top-down signaling (***Nickel et al., 2020***). Nevertheless, the focus of these studies was on generic interactions (i.e. top-down vs. bottom-up) processes without directly inducing prediction errors as in a cued pain paradigm employed in our study.

In the flexible routing model proposed by ***Ploner et al., 2017***, pain is seen as driven by contextual processes, such as expectations, which is associated with alpha/beta oscillations and alpha/beta synchrony across brain areas. Previous studies have started to examine the spectral properties of mechanisms related to generative models of pain perception. In particular, a previous MEG study reported that alpha suppression in the anterior insula is related mainly to pain expectation in a paradigm in which painful stimuli were interleaved with non-painful stimuli (***Franciotti et al., 2009***). This was interpreted as a preparatory mechanism for an upcoming painful stimulus. In a related study, alpha desynchronization in the context of predictable painful stimuli has been discussed as a possible neural correlate of attentional preparatory processes (***Babiloni et al., 2003***).

Expectation is also a crucial ingredient of placebo analgesia and nocebo hyperalgesia. A previous study reported that resting-state alpha band activity was also linked to the expectation of pain modulation (analgesia) in a placebo paradigm (***Huneke et al., 2013***). With respect to negative expectations, it has been shown that pain modulation due to nocebo expectation is associated with enhanced alpha activity (***Albu and Meagher, 2016***). Our findings are in line with these results indicating an important role of low frequency activity in mediating expectation effects in a pain network underlying a generative model for pain perception.

In contrast to prediction error effects in the visual (***Bauer et al., 2014***; ***van Pelt et al., 2016***) and auditory (***Edwards et al., 2005***; ***Parras et al., 2017***) domains, we observed a negative modulation of gamma activity by absolute prediction errors. However, it should be noted that opposite effects have been observed in other cognitive domains. For instance, increased gamma power has been associated with successful matching (i.e. the absence of a prediction error) between external input and internal representation (***Herrmann et al., 2004a***; ***Osipova et al., 2006***; ***Wang et al., 2018***). In particular, gamma band responses have been explained in terms of the match between bottom-up and top-down information (***Herrmann et al., 2004b***). One example is the observation of increased gamma activity with a higher so-called *cloze probability* in sentence-level language comprehension (***Hald et al., 2006***; ***Obleser and Kotz, 2011***; ***Wang et al., 2012***; ***Wang et al., 2018***; ***Molinaro et al., 2013***). It has been shown that a critical word that is semantically predictable by the preceding sentence (so-called high cloze probability) induces a larger gamma response than words which are semantically incongruent (i.e. unpredicted; low cloze probability) (***Wang et al., 2018***).

## Pain vs. thermoception

In the present study, the lowest stimulus intensity was often not perceived as painful but as hot. In general, stimulus properties were chosen to be comparable to a previous fMRI study which showed fMRI signals related to prediction errors (***Fazeli and Büchel, 2018***). However, even though the lowest stimulus intensity (42˚C) was above the threshold of nociceptors (***Treede et al., 1998***), the subjective experience of the lowest pain stimuli was often rated as neutral. Therefore, we performed an additional analysis (reduced pain model) only comprising clearly painful stimuli (46˚C and 48˚C) to

more specifically address expectations and predictions errors in pain. The analyses of the behavioral data revealed similar results. Both models showed a highly significant effect of stimulus intensity and expectation on perceived stimulus intensity. In addition, the reduced pain model showed a significant prediction error effect, which was formally not observed in the full model. However, it is important to note that this difference should not be overinterpreted, as the p-value for the prediction error effect of the full model was at a trend level (p=0.1). Importantly, the negative representation of prediction errors in the gamma band was evident in both, the reduced and the full model.

### Limitations

To unravel the temporal aspects of expectations and prediction errors, this study has been designed in close analogy to a previous fMRI study and we decided to use the same experimental paradigm (***Fazeli and Büchel, 2018***). We therefore decided to also keep the sample characteristics similar and restricted the sample to male participants, which means that we cannot generalize our results to the population. However, our study agrees with the findings of a previous study using a similar design (***Geuter et al., 2017***) which tested male and female participants. Future studies should investigate samples including female participants. This would also allow to investigate sex effects with respect to expectation and prediction error effects in pain.

To minimize motor responses and speed up the rating procedure, we used a four-button device to directly assess stimulus intensity (in contrast to using two buttons to move a slider on a VAS), thus being limited to a coarse rating scale of four levels, where one was labeled as 'neutral' and four was labeled as 'very strong'. This allows to accommodate more trials but is not ideal to assess fine-grained differences, specifically to differentiate between non-painful and painful stimulation, as level 1 would represent 0–25 on a 0–100 VAS. Future research could use conventional 0–100 VAS to assess stimulus intensity on a finer scale.

For reasons of comparability to a previous fMRI study, we employed three different temperatures for all volunteers. Alternatively, we could have defined three levels of pain based on individual calibration of heat stimuli (***Taesler and Rose, 2017***; ***Grahl et al., 2018***; ***Horing et al., 2019***; ***Zhang et al., 2020***; ***Feldhaus et al., 2021***). Such a procedure could have avoided trials where no pain was subjectively perceived. On the other hand, such an approach also carries the risk that subjective ratings during the calibration process do not truly reflect pain and can lead to errors (especially if ratings are too low) which then affect the entire experiment. However, to address this shortcoming, we performed an additional analysis, which only included painful stimulus intensities.

### Summary

Our data show that key variables required for pain perception and thermoception in the context of a generative model are correlated with distinct oscillatory profiles in the brain. Furthermore, each oscillatory frequency band was correlated with a distinct variable such as expectation and prediction errors. These mechanistic insights could be very helpful in patients with acute and more importantly in patients with chronic pain, where expectations have been shown to play a critical role in pain persistence.

## Acknowledgements

We would like to thank Markus Ploner for comments on an earlier version of this manuscript. We would also like to thank Matthias Kerkemeyer for his help during data collection. CB is supported by DFG SFB 289 project A02 and ERC-AdG-883892-PainPersist. MR is supported by DFG SFB 289 project A03 and DFG SFB TR 169 project B3. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 422744262–TRR 289.

## Additional information

### Funding


| Funder | Grant reference number | Author |
|---|---|---|
| Deutsche Forschungsgemeinschaft | 422744262–TRR 289 project A02 | Christian Büchel |

| | | |
|---|---|---|
| Deutsche Forschungsgemeinschaft | DFG SFB TR 169 project B3 | Michael Rose |
| H2020 European Research Council | ERC-AdG-883892-PainPersist | Christian Büchel |
| Deutsche Forschungsgemeinschaft | 422744262–TRR 289 project A03 | Michael Rose |

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.


### Author contributions

Andreas Strube, Conceptualization, Data curation, Software, Formal analysis, Investigation, Visualization, Methodology, Writing - original draft, Project administration, Writing - review and editing; Michael Rose, Conceptualization, Resources, Software, Methodology, Writing - review and editing; Sepideh Fazeli, Conceptualization, Software, Supervision; Christian Büchel, Conceptualization, Resources, Formal analysis, Supervision, Funding acquisition, Validation, Visualization, Methodology, Project administration, Writing - review and editing


### Author ORCIDs

Andreas Strube https://orcid.org/0000-0002-6545-0366
Christian Büchel https://orcid.org/0000-0003-1965-906X


### Ethics

Human subjects: All volunteers gave their informed consent. The study was approved by the Ethics board of the Hamburg Medical Association (PV4745).

### Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.62809.sa1
Author response https://doi.org/10.7554/eLife.62809.sa2

## Additional files

### Supplementary files

- Transparent reporting form

### Data availability

Data for this study are available on https://osf.io/f2mua/.

The following dataset was generated:

| Author(s) | Year | Dataset title | Dataset URL | Database and Identifier |
|---|---|---|---|---|
| Strube A, Rose M, Fazeli S, Büchel C | 2020 | The temporal and spectral characteristics of expectations and prediction errors in pain and thermoception | https://osf.io/f2mua/ | Open Science Framework, F2MUA |

## References


**Albu S**, Meagher MW. 2016. Expectation of nocebo hyperalgesia affects EEG alpha-activity. *International Journal of Psychophysiology* **109**:147–152. DOI: https://doi.org/10.1016/j.ijpsycho.2016.08.009, PMID: 27562424

**Anchisi D**, Zanon M. 2015. A bayesian perspective on sensory and cognitive integration in pain perception and placebo analgesia. *PLOS ONE* **10**:e0117270. DOI: https://doi.org/10.1371/journal.pone.0117270, PMID: 25664586

**Arnal LH**, Giraud AL. 2012. Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences* **16**:390–398. DOI: https://doi.org/10.1016/j.tics.2012.05.003, PMID: 22682813

**Atlas LY**, Bolger N, Lindquist MA, Wager TD. 2010. Brain mediators of predictive cue effects on perceived pain. *Journal of Neuroscience* **30**:12964–12977. DOI: https://doi.org/10.1523/JNEUROSCI.0057-10.2010, PMID: 20881115

**Atlas LY**, Wager TD. 2012. How expectations shape pain. *Neuroscience Letters* **520**:140–148. DOI: https://doi.org/10.1016/j.neulet.2012.03.039, PMID: 22465136

**Babiloni C**, Brancucci A, Babiloni F, Capotosto P, Carducci F, Cincotti F, Arendt-Nielsen L, Chen ACN, Rossini PM. 2003. Anticipatory cortical responses during the expectancy of a predictable painful stimulation A high-resolution electroencephalography study. *European Journal of Neuroscience* **18**:1692–1700. DOI: https://doi.org/10.1046/j.1460-9568.2003.02851.x

**Bastos AM**, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ. 2012. Canonical microcircuits for predictive coding. *Neuron* **76**:695–711. DOI: https://doi.org/10.1016/j.neuron.2012.10.038, PMID: 23177956

**Bauer M**, Oostenveld R, Peeters M, Fries P. 2006. Tactile spatial attention enhances gamma-band activity in somatosensory cortex and reduces low-frequency activity in parieto-occipital Areas. *Journal of Neuroscience* **26**:490–501. DOI: https://doi.org/10.1523/JNEUROSCI.5228-04.2006, PMID: 16407546

**Bauer M**, Stenner MP, Friston KJ, Dolan RJ. 2014. Attentional modulation of alpha/beta and gamma oscillations reflect functionally distinct processes. *Journal of Neuroscience* **34**:16117–16125. DOI: https://doi.org/10.1523/JNEUROSCI.3474-13.2014, PMID: 25429152

**Bingel U**, Lorenz J, Schoell E, Weiller C, Büchel C. 2006. Mechanisms of placebo analgesia: racc recruitment of a subcortical antinociceptive network. *Pain* **120**:8–15. DOI: https://doi.org/10.1016/j.pain.2005.08.027, PMID: 16364549

**Büchel C**, Geuter S, Sprenger C, Eippert F. 2014. Placebo analgesia: a predictive coding perspective. *Neuron* **81**:1223–1239. DOI: https://doi.org/10.1016/j.neuron.2014.02.042, PMID: 24656247

**Clark A**. 2013. Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* **36**:181–204. DOI: https://doi.org/10.1017/S0140525X12000477

**Colloca L**, Benedetti F. 2005. Placebos and painkillers: is mind as real as matter? *Nature Reviews Neuroscience* **6**:545–552. DOI: https://doi.org/10.1038/nrn1705, PMID: 15995725

**Debener S**, Herrmann CS, Kranczioch C, Gembris D, Engel AK. 2003. Top-down attentional processing enhances auditory evoked gamma band activity. *NeuroReport* **14**:683–686. DOI: https://doi.org/10.1097/00001756-200304150-00005, PMID: 12692463

**Edwards E**, Soltani M, Deouell LY, Berger MS, Knight RT. 2005. High gamma activity in response to deviant auditory stimuli recorded directly from human cortex. *Journal of Neurophysiology* **94**:4269–4280. DOI: https://doi.org/10.1152/jn.00324.2005, PMID: 16093343

**Egner T**, Monti JM, Summerfield C. 2010. Expectation and surprise determine neural population responses in the ventral visual stream. *Journal of Neuroscience* **30**:16601–16608. DOI: https://doi.org/10.1523/JNEUROSCI.2770-10.2010, PMID: 21147999

**Ernst MO**, Banks MS. 2002. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**:429–433. DOI: https://doi.org/10.1038/415429a, PMID: 11807554

**Fazeli S**, Büchel C. 2018. Pain-Related expectation and prediction error signals in the anterior insula are not related to aversiveness. *The Journal of Neuroscience* **38**:6461–6474. DOI: https://doi.org/10.1523/JNEUROSCI.0671-18.2018, PMID: 29934355

**Feldhaus MH**, Horing B, Sprenger C, Büchel C. 2021. Association of nocebo hyperalgesia and basic somatosensory characteristics in a large cohort. *Scientific Reports* **11**:762. DOI: https://doi.org/10.1038/s41598-020-80386-y, PMID: 33436821

**Franciotti R**, Ciancetta L, Della Penna S, Belardinelli P, Pizzella V, Romani GL. 2009. Modulation of alpha oscillations in insular cortex reflects the threat of painful stimuli. *NeuroImage* **46**:1082–1090. DOI: https://doi.org/10.1016/j.neuroimage.2009.03.034, PMID: 19327401

**Friston K**. 2010. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience* **11**:127–138. DOI: https://doi.org/10.1038/nrn2787, PMID: 20068583

**Geuter S**, Boll S, Eippert F, Büchel C. 2017. Functional dissociation of stimulus intensity encoding and predictive coding of pain in the insula. *eLife* **6**:e24770. DOI: https://doi.org/10.7554/eLife.24770, PMID: 28524817

**Grahl A**, Onat S, Büchel C. 2018. The periaqueductal gray and bayesian integration in placebo analgesia. *eLife* **7**:e32930. DOI: https://doi.org/10.7554/eLife.32930, PMID: 29555019

**Grandchamp R**, Delorme A. 2011. Single-trial normalization for event-related spectral decomposition reduces sensitivity to noisy trials. *Frontiers in Psychology* **2**:236. DOI: https://doi.org/10.3389/fpsyg.2011.00236, PMID: 21994498

**Gross J**, Schnitzler A, Timmermann L, Ploner M. 2007. Gamma oscillations in human primary somatosensory cortex reflect pain perception. *PLOS Biology* **5**:e133. DOI: https://doi.org/10.1371/journal.pbio.0050133, PMID: 17456008

**Gruber T**, Müller MM, Keil A, Elbert T. 1999. Selective visual-spatial attention alters induced gamma band responses in the human EEG. *Clinical Neurophysiology* **110**:2074–2085. DOI: https://doi.org/10.1016/S1388-2457(99)00176-5, PMID: 10616112

**Hald LA**, Bastiaansen MC, Hagoort P. 2006. EEG theta and gamma responses to semantic violations in online sentence processing. *Brain and Language* **96**:90–105. DOI: https://doi.org/10.1016/j.bandl.2005.06.007, PMID: 16083953

**Hauck M**, Lorenz J, Engel AK. 2007. Attention to painful stimulation enhances gamma-band activity and synchronization in human sensorimotor cortex. *Journal of Neuroscience* **27**:9270–9277. DOI: https://doi.org/10.1523/JNEUROSCI.2283-07.2007, PMID: 17728441

**Hauck M**, Domnick C, Lorenz J, Gerloff C, Engel AK. 2015. Top-down and bottom-up modulation of pain-induced oscillations. *Frontiers in Human Neuroscience* **9**:375. DOI: https://doi.org/10.3389/fnhum.2015.00375, PMID: 26190991

**Herrmann CS**, Lenz D, Junge S, Busch NA, Maess B. 2004a. Memory-matches evoke human gamma-responses. *BMC Neuroscience* **5**:13. DOI: https://doi.org/10.1186/1471-2202-5-13, PMID: 15084225

**Herrmann CS**, Munk MH, Engel AK. 2004b. Cognitive functions of gamma-band activity: memory match and utilization. *Trends in Cognitive Sciences* **8**:347–355. DOI: https://doi.org/10.1016/j.tics.2004.06.006, PMID: 15335461

**Horing B**, Sprenger C, Büchel C. 2019. The parietal operculum preferentially encodes heat pain and not salience. *PLOS Biology* **17**:e3000205. DOI: https://doi.org/10.1371/journal.pbio.3000205, PMID: 31404058

**Hu L**, Peng W, Valentini E, Zhang Z, Hu Y. 2013. Functional features of nociceptive-induced suppression of alpha band electroencephalographic oscillations. *The Journal of Pain* **14**:89–99. DOI: https://doi.org/10.1016/j.jpain.2012.10.008, PMID: 23273836

**Huang Y**, Rao RPN, Rpn R. 2011. Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science* **2**:580–593. DOI: https://doi.org/10.1002/wcs.142

**Huneke NT**, Brown CA, Burford E, Watson A, Trujillo-Barreto NJ, El-Deredy W, Jones AK. 2013. Experimental placebo analgesia changes resting-state alpha oscillations. *PLOS ONE* **8**:e78278. DOI: https://doi.org/10.1371/journal.pone.0078278, PMID: 24147129

**Jung TP**, Makeig S, Humphries C, Lee TW, McKeown MJ, Iragui V, Sejnowski TJ. 2000. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* **37**:163–178. DOI: https://doi.org/10.1111/1469-8986.3720163, PMID: 10731767

**Klem GH**, Lüders HO, Jasper HH, Elger C. 1999. The ten-twenty electrode system of the international federation the international federation of clinical neurophysiology. *Electroencephalography and Clinical Neurophysiology. Supplement* **52**:3–6. PMID: 10590970

**Knill DC**, Pouget A. 2004. The bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences* **27**:712–719. DOI: https://doi.org/10.1016/j.tins.2004.10.007, PMID: 15541511

**Koyama T**, McHaffie JG, Laurienti PJ, Coghill RC. 2005. The subjective experience of pain: where expectations become reality. *PNAS* **102**:12950–12955. DOI: https://doi.org/10.1073/pnas.0408576102, PMID: 16150703

**Lang PJ**, Bradley MM, Cuthbert BN. 2008. *International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual. Technical Report a-8*: University of Florida. https://imotions.com/blog/iaps-international-affective-picture-system/.

**Liberati G**, Klöcker A, Algoet M, Mulders D, Maia Safronova M, Ferrao Santos S, Ribeiro Vaz JG, Raftopoulos C, Mouraux A. 2018. Gamma-Band oscillations preferential for nociception can be recorded in the human insula. *Cerebral Cortex* **28**:3650–3664. DOI: https://doi.org/10.1093/cercor/bhx237, PMID: 29028955

**Lorenz J**, Garcia-Larrea L. 2003. Contribution of attentional and cognitive factors to laser evoked brain potentials. *Neurophysiologie Clinique/Clinical Neurophysiology* **33**:293–301. DOI: https://doi.org/10.1016/j.neucli.2003.10.004

**Makeig S**, Bell A, Jung T-P, Sejnowski T. 1996. Independent component analysis of electroencephalographic data. NIPS'95: Proceedings of the 8th International Conference on Neural Information Processing Systems 145–151. DOI: https://doi.org/10.5555/2998828.2998849

**Maris E**, Oostenveld R. 2007. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods* **164**:177–190. DOI: https://doi.org/10.1016/j.jneumeth.2007.03.024, PMID: 17517438

**May ES**, Butz M, Kahlbrock N, Hoogenboom N, Brenner M, Schnitzler A. 2012. Pre- and post-stimulus alpha activity shows differential modulation with spatial attention during the processing of pain. *NeuroImage* **62**:1965–1974. DOI: https://doi.org/10.1016/j.neuroimage.2012.05.071, PMID: 22659486

**Molinaro N**, Barraza P, Carreiras M. 2013. Long-range neural synchronization supports fast and efficient reading: eeg correlates of processing expected words in sentences. *NeuroImage* **72**:120–132. DOI: https://doi.org/10.1016/j.neuroimage.2013.01.031, PMID: 23357072

**Mouraux A**, Guérit JM, Plaghki L. 2003. Non-phase locked electroencephalogram (EEG) responses to CO2 laser skin stimulations may reflect central interactions between A partial partial differential- and C-fibre afferent volleys. *Clinical Neurophysiology* **114**:710–722. DOI: https://doi.org/10.1016/S1388-2457(03)00027-0, PMID: 12686279

**Nickel MM**, Ta Dinh S, May ES, Tiemann L, Hohn VD, Gross J, Ploner M. 2020. Neural oscillations and connectivity characterizing the state of tonic experimental pain in humans. *Human Brain Mapping* **41**:17–29. DOI: https://doi.org/10.1002/hbm.24784, PMID: 31498948

**Obleser J**, Kotz SA. 2011. Multiple brain signatures of integration in the comprehension of degraded speech. *NeuroImage* **55**:713–723. DOI: https://doi.org/10.1016/j.neuroimage.2010.12.020, PMID: 21172443

**Ongaro G**, Kaptchuk TJ. 2019. Symptom perception, placebo effects, and the bayesian brain. *Pain* **160**:1–4. DOI: https://doi.org/10.1097/j.pain.0000000000001367, PMID: 30086114

**Oostenveld R**, Fries P, Maris E, Schoffelen JM. 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience* **2011**:1–9. DOI: https://doi.org/10.1155/2011/156869, PMID: 21253357

**Osipova D**, Takashima A, Oostenveld R, Fernández G, Maris E, Jensen O. 2006. Theta and gamma oscillations predict encoding and retrieval of declarative memory. *Journal of Neuroscience* **26**:7523–7531. DOI: https://doi.org/10.1523/JNEUROSCI.1948-06.2006, PMID: 16837600

**Parras GG**, Nieto-Diego J, Carbajal GV, Valdés-Baizabal C, Escera C, Malmierca MS. 2017. Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nature Communications* **8**:2148. DOI: https://doi.org/10.1038/s41467-017-02038-6, PMID: 29247159

**Petrovic P**, Kalso E, Petersson KM, Ingvar M. 2002. Placebo and opioid analgesia– imaging a shared neuronal network. *Science* **295**:1737–1740. DOI: https://doi.org/10.1126/science.1067176, PMID: 11834781

**Ploghaus A**, Tracey I, Gati JS, Clare S, Menon RS, Matthews PM, Rawlins JN. 1999. Dissociating pain from its anticipation in the human brain. *Science* **284**:1979–1981. DOI: https://doi.org/10.1126/science.284.5422.1979, PMID: 10373114

**Ploner M**, Gross J, Timmermann L, Pollok B, Schnitzler A. 2006. Pain suppresses spontaneous brain rhythms. *Cerebral Cortex* **16**:537–540. DOI: https://doi.org/10.1093/cercor/bhj001, PMID: 16033927

**Ploner M**, Sorg C, Gross J. 2017. Brain rhythms of pain. *Trends in Cognitive Sciences* **21**:100–110. DOI: https://doi.org/10.1016/j.tics.2016.12.001, PMID: 28025007

**Rossiter HE**, Worthen SF, Witton C, Hall SD, Furlong PL. 2013. Gamma oscillatory amplitude encodes stimulus intensity in primary somatosensory cortex. *Frontiers in Human Neuroscience* **7**:362. DOI: https://doi.org/10.3389/fnhum.2013.00362, PMID: 23874282

**Schulz E**, May ES, Postorino M, Tiemann L, Nickel MM, Witkovsky V, Schmidt P, Gross J, Ploner M. 2015. Prefrontal gamma oscillations encode tonic pain in humans. *Cerebral Cortex* **25**:4407–4414. DOI: https://doi.org/10.1093/cercor/bhv043, PMID: 25754338

**Song Y**, Yao M, Kemprecos H, Á B, Xiao Z, Zhang Q, Singh A, Wang J, Chen ZS. 2019. Predictive coding models for pain perception. *Neuroscience* **17**:00780-x. DOI: https://doi.org/10.1007/s10827-021-00780-x

**Summerfield C**, de Lange FP. 2014. Expectation in perceptual decision making: neural and computational mechanisms. *Nature Reviews Neuroscience* **15**:745–756. DOI: https://doi.org/10.1038/nrn3838, PMID: 25315388

**Taesler P**, Rose M. 2016. Prestimulus theta oscillations and connectivity modulate pain perception. *The Journal of Neuroscience* **36**:5026–5033. DOI: https://doi.org/10.1523/JNEUROSCI.3325-15.2016

**Taesler P**, Rose M. 2017. Psychophysically-anchored, robust thresholding in studying Pain-related lateralization of oscillatory prestimulus activity. *Journal of Visualized Experiments* **2017**:55228. DOI: https://doi.org/10.3791/55228

**Tiemann L**, Schulz E, Gross J, Ploner M. 2010. Gamma oscillations as a neuronal correlate of the attentional effects of pain. *Pain* **150**:302–308. DOI: https://doi.org/10.1016/j.pain.2010.05.014, PMID: 20558000

**Tiemann L**, May ES, Postorino M, Schulz E, Nickel MM, Bingel U, Ploner M. 2015. Differential neurophysiological correlates of bottom-up and top-down modulations of pain. *Pain* **156**:289–296. DOI: https://doi.org/10.1097/01.j.pain.0000460309.94442.44, PMID: 25599450

**Tiitinen H**, Sinkkonen J, Reinikainen K, Alho K, Lavikainen J, Näätänen R. 1993. Selective attention enhances the auditory 40-Hz transient response in humans. *Nature* **364**:59–60. DOI: https://doi.org/10.1038/364059a0, PMID: 8316297

**Todorovic A**, van Ede F, Maris E, de Lange FP. 2011. Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. *Journal of Neuroscience* **31**:9118–9123. DOI: https://doi.org/10.1523/JNEUROSCI.1425-11.2011

**Tracey I**, Mantyh PW. 2007. The cerebral signature for pain perception and its modulation. *Neuron* **55**:377–391. DOI: https://doi.org/10.1016/j.neuron.2007.07.012, PMID: 17678852

**Treede RD**, Meyer RA, Campbell JN. 1998. Myelinated mechanically insensitive afferents from monkey hairy skin: heat-response properties. *Journal of Neurophysiology* **80**:1082–1093. DOI: https://doi.org/10.1152/jn.1998.80.3.1082, PMID: 9744923

**Tu Y**, Zhang Z, Tan A, Peng W, Hung YS, Moayedi M, Iannetti GD, Hu L. 2016. Alpha and gamma oscillation amplitudes synergistically predict the perception of forthcoming nociceptive stimuli. *Human Brain Mapping* **37**:501–514. DOI: https://doi.org/10.1002/hbm.23048, PMID: 26523484

**van Pelt S**, Heil L, Kwisthout J, Ondobaka S, van Rooij I, Bekkering H. 2016. Beta- and gamma-band activity reflect predictive coding in the processing of causal events. *Social Cognitive and Affective Neuroscience* **11**:973–980. DOI: https://doi.org/10.1093/scan/nsw017, PMID: 26873806

**Wager TD**, Rilling JK, Smith EE, Sokolik A, Casey KL, Davidson RJ, Kosslyn SM, Rose RM, Cohen JD. 2004. Placebo-induced changes in FMRI in the anticipation and experience of pain. *Science* **303**:1162–1167. DOI: https://doi.org/10.1126/science.1093065, PMID: 14976306

**Wang L**, Zhu Z, Bastiaansen M. 2012. Integration or predictability? A further specification of the functional role of gamma oscillations in language comprehension. *Frontiers in Psychology* **3**:00187. DOI: https://doi.org/10.3389/fpsyg.2012.00187

**Wang L**, Hagoort P, Jensen O. 2018. Gamma oscillatory activity related to language prediction. *Journal of Cognitive Neuroscience* **30**:1075–1085. DOI: https://doi.org/10.1162/jocn_a_01275, PMID: 29708821

**Wiech K**, Ploner M, Tracey I. 2008. Neurocognitive aspects of pain perception. *Trends in Cognitive Sciences* **12**:306–313. DOI: https://doi.org/10.1016/j.tics.2008.05.005, PMID: 18606561

**Wiech K**. 2016. Deconstructing the sensation of pain: the influence of cognitive processes on pain perception. *Science* **354**:584–587. DOI: https://doi.org/10.1126/science.aaf8934, PMID: 27811269

**Wilkinson D**, Halligan P. 2004. The relevance of behavioural measures for functional-imaging studies of cognition. *Nature Reviews Neuroscience* **5**:67–73. DOI: https://doi.org/10.1038/nrn1302, PMID: 14708005

**Xiao Z**, Martinez E, Kulkarni PM, Zhang Q, Hou Q, Rosenberg D, Talay R, Shalot L, Zhou H, Wang J, Chen ZS. 2019. Cortical pain processing in the rat anterior cingulate cortex and primary somatosensory cortex. *Frontiers in Cellular Neuroscience* **13**:165. DOI: https://doi.org/10.3389/fncel.2019.00165, PMID: 31105532

**Zhang ZG**, Hu L, Hung YS, Mouraux A, Iannetti GD. 2012. Gamma-band oscillations in the primary somatosensory cortex–a direct and obligatory correlate of subjective pain intensity. *Journal of Neuroscience* **32**:7429–7438. DOI: https://doi.org/10.1523/JNEUROSCI.5877-11.2012, PMID: 22649223

**Zhang S**, Yoshida W, Mano H, Yanagisawa T, Mancini F, Shibata K, Kawato M, Seymour B. 2020. Pain control by Co-adaptive learning in a Brain-Machine interface. *Current Biology* **30**:3935–3944. DOI: https://doi.org/10.1016/j.cub.2020.07.066

## 15.2 Study 2: Alpha-to-Beta- and Gamma-Band Activity Reflect Predictive Coding in Affective Visual Processing

Strube, A., Rose, M., Fazeli, S., & Büchel, C. (2021). Alpha-to-Beta and Gamma-Band Activity Reflect Predictive Coding in Affective Visual Processing. Scientific Reports, 11(1), 1-15.

scientific reports

Check for updates

OPEN

# Alpha-to-beta- and gamma-band activity reflect predictive coding in affective visual processing

Andreas Strube✉, Michael Rose, Sepideh Fazeli & Christian Büchel

**Processing of negative affective pictures typically leads to desynchronization of alpha-to-beta frequencies (ERD) and synchronization of gamma frequencies (ERS). Given that in predictive coding higher frequencies have been associated with prediction errors, while lower frequencies have been linked to expectations, we tested the hypothesis that alpha-to-beta ERD and gamma ERS induced by aversive pictures are associated with expectations and prediction errors, respectively. We recorded EEG while volunteers were involved in a probabilistically cued affective picture task using three different negative valences to produce expectations and prediction errors. Our data show that alpha-to-beta band activity after stimulus presentation was related to the expected valence of the stimulus as predicted by a cue. The absolute mismatch of the expected and actual valence, which denotes an absolute prediction error was related to increases in alpha, beta and gamma band activity. This demonstrates that top-down predictions and bottom-up prediction errors are represented in typical spectral patterns associated with affective picture processing. This study provides direct experimental evidence that negative affective picture processing can be described by neuronal predictive coding computations.**

People see hundreds of unfamiliar faces in daily life, while seeing famous faces is very rare and surprising and—in terms of predictive coding—unexpected, leading to a large prediction error. Utilizing time–frequency analysis of brain data, it has been shown for example that famous faces elicit larger gamma responses as compared to unfamiliar faces[1,2].

Predictive coding of perception assumes that neuronal circuits implement perception and learning by constantly matching incoming sensory data with the top-down predictions of an internal or generative model[3–5]. Consequently, a system can refine models with better predictions by minimizing prediction errors regarding the sensory environment, leading to a more efficient encoding of information[6].

The Free Energy principle including aspects of predictive coding specifically posits the minimization of "free energy" (and thus, indirectly prediction errors) as a mechanism to ensure that agents spend most of their time in a small number of valuable (i.e. positive) and expected states[6]. With regards to affective stimuli, this agrees with findings showing that visual stimuli with a negative valence (i.e. a negative and thus unexpected state) produce larger gamma responses than neutral and positive visual stimuli[7–16]. Results interpreting the effects of negative valence in the gamma band could be associated with the surprise (i.e. general low probability of a negative encounter) that negative stimuli entail. However, in most studies this cannot be disentangled from the valence as the prediction error associated with a negative stimulus per se cannot be disentangled from the prediction error in the individual experimental setting. To achieve this, additional prediction errors have to be introduced experimentally.

Within the framework of predictive coding, lower frequency oscillatory alpha-to-beta band activity has been linked to top-down predictive signals and higher frequency gamma band activity to bottom-up prediction errors[4,17,18]. Comparably, cortical dynamics induced by emotional picture processing comprise event-related desynchronization (ERD) in the alpha-to-beta band[19–28] and event-related synchronization (ERS) in the gamma band[9,12,14,22,26,29–33].

Alpha ERD (a decrease in power in the ~ 8–12 Hz range) following affective images is smaller when the image is anticipated, and the tendency is more prominent for images bearing negative emotional valence[34]. This might be interpreted as differences in the encoding of expectation signals in a predictive coding framework, where expectation signals manifest as increases in low frequency (alpha-to-beta) activity. In this context, predictive

Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany. ✉email: a.strube@uke.de

coding suggests that feed-forward prediction errors reflect the difference between top-down expectation signals (e.g. pre-activated neuronal units based on the anticipation of threat) and actual stimulus input[17].

In the context of affective picture processing, it is interesting to note that participants with dysphoria elicit a smaller alpha ERD in response to pleasant pictures, but not to unpleasant pictures[24]. In the context of predictive coding, affective disorders (such as major depression) have been linked with bottom-up deficits in predictive processing and increased precision of negative prior beliefs[35]. The depressed phenotype may emerge from a collection of depressive beliefs associated with the causal structure of the world[36]. As a consequence, treating depression could be conceptualized as equipping the brain with the resources to modify its internal model of the world[37]. Hence, treatment of depression would be associated with brain's relevant statistical structures becoming "less pessimistic"[36]. Thus, the predictive coding model of emotional states associated with affective disorders might be of particular interest for mechanistic insights in depression, as it represents such an internal model with potentially pathological variations in its statistical structure[35–39].

In summary, we hypothesize that alpha-to-beta ERD and gamma ERS typically found in responses to negative affective stimuli are actually signals related to predictive coding. This posits that alpha-to-beta ERD responses should be modulated by expectations, whereas gamma ERS responses should be modulated by prediction errors or surprise.

Consequently, we conducted a cue-stimulus paradigm to unravel predictive coding dynamics in affective picture processing. We specifically introduced prediction errors experimentally by presenting stimuli in two different modalities, pain and vision (i.e. affective pictures). In the affective picture part presented here, we presented emotionally negative stimuli and manipulated the degree of negative valence. Participants were asked to rate the valence of the content on a four-point rating scale. We expected the anticipated degree of the aversive content to be related to alpha-to-beta ERD. If surprise is a main driving factor of gamma ERS (as derived from a predictive coding perspective), we expected an increase of gamma power when there was a mismatch between the anticipated degree of aversion and the actual aversive quality of the picture. If the negative valence or aversive quality is contributing to the gamma ERS effect, we expected an increase of gamma power with higher aversion regardless of the anticipated degree of aversion. Finally, based on hypotheses regarding a negative valence associated with prediction errors[40], we expected that a greater mismatch between predicted and actual valence elicits larger valence ratings.

## Methods

**Participants.** We investigated 35 healthy male participants (mean 26, range 18–37 years), who were paid as compensation for their participation. Applicants were excluded if one of the following exclusion criteria applied: neurological, psychiatric, dermatological diseases, pain conditions, current medication, or substance abuse. All volunteers gave their informed consent. The study was approved by the Ethics board of the Hamburg Medical Association. Data from six participants had to be excluded from the final EEG data analysis due to technical issues during the EEG recording (i.e. the data of the excluded participants were contaminated with excessive muscle and/or technical artifacts) leaving a final sample of 29 participants.

**Stimuli and task.** Stimulus properties were chosen to be identical to a previous fMRI study of predictive coding where both expectation and absolute prediction error effects were observed in pain[41].

Aversive pictures were chosen from the International Affective Picture System (IAPS)[42] database at three different levels of valence. The images presented during the EEG experiment had three levels of valence of which the low valence category had valence values of $2.02 \pm 0.05$ (mean ± standard error), the medium valence category had valence values of $4.06 \pm 0.02$ (mean ± standard error) and the high valence category had valence values of $5.23 \pm 0.01$ (mean ± standard error). The pain part of this data will not be described here, but has been reported in Strube et al., 2021[43].

Prior to each picture or heat stimulus, a visual cue was presented. The color of the cue (triangle, visual angle of each side: 0.96°) indicated (probabilistically) the modality of the stimulus (orange for picture and blue for heat). A white digit depicted inside of each triangle indicated (probabilistically) the intensity of the subsequent stimulus (1, 2 and 3 for low, medium and high valence). During the whole trial, a centered fixation cross (visual angle: 0.24°) was presented on the screen.

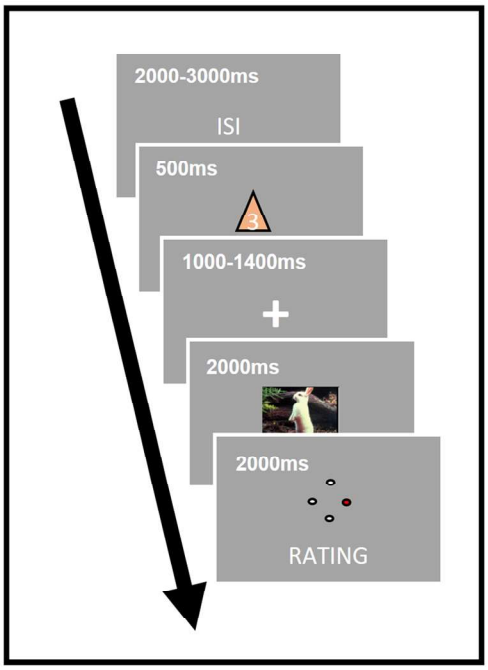
Each trial began with the presentation of the cue for 500 ms as an indicator for the modality and intensity of the subsequently presented stimulus. The modality (i.e. pain or picture) was correctly cued in 70% of all trials by the color of the triangle. In 60% of all trials, the stimulus intensity was correctly indicated by the digit within the triangle (see Fig. 1b for an overview of all cue contingencies).

Before the presentation of the stimulus, there was a blank period with a variable duration between 1000 and 1400 ms. The visual (or thermal) stimulus was presented for a duration of two seconds. The visual stimulus (horizontal visual angle of 3.8°; vertical visual angle of 2.4°) was centered on the screen and allowed the participant to perceive it without eye movements. After the termination of the stimulus, subjects were asked to rate the aversiveness of the stimulus on a four point rating scale, where 1 was labeled as "neutral" and 4 was labeled as "very strong". Ratings were performed using a response box operated with the right hand (see Fig. 1a for a visualization of the trial structure).

In addition, four catch trials were included in each block. Subjects were asked to report the preceding cue in terms of their information content of the modality and intensity within 8 s and no stimulation was given in these trials.

Trials were presented in four blocks. Each block consisted of 126 trials and four catch trials and lasted about 15 min. The trial order within each block was pseudorandomized. The order of blocks was randomized across subjects. The whole EEG experiment including preparation and instructions lasted for about three hours.

a)




b)

| | Low Valence Picture | Medium Valence Picture | High Valence Picture | Low Pain | Medium Pain | High Pain |
|---|---|---|---|---|---|---|
| 1 | 7.0% | 2.3% | 2.3% | 3.0% | 1.0% | 1.0% |
| 2 | 2.3% | 7.0% | 2.3% | 1.0% | 3.0% | 1.0% |
| 3 | 2.3% | 2.3% | 7.0% | 1.0% | 1.0% | 3.0% |
| 1 | 3.0% | 1.0% | 1.0% | 7.0% | 2.3% | 2.3% |
| 2 | 1.0% | 3.0% | 1.0% | 2.3% | 7.0% | 2.3% |
| 3 | 1.0% | 1.0% | 3.0% | 2.3% | 2.3% | 7.0% |

c)

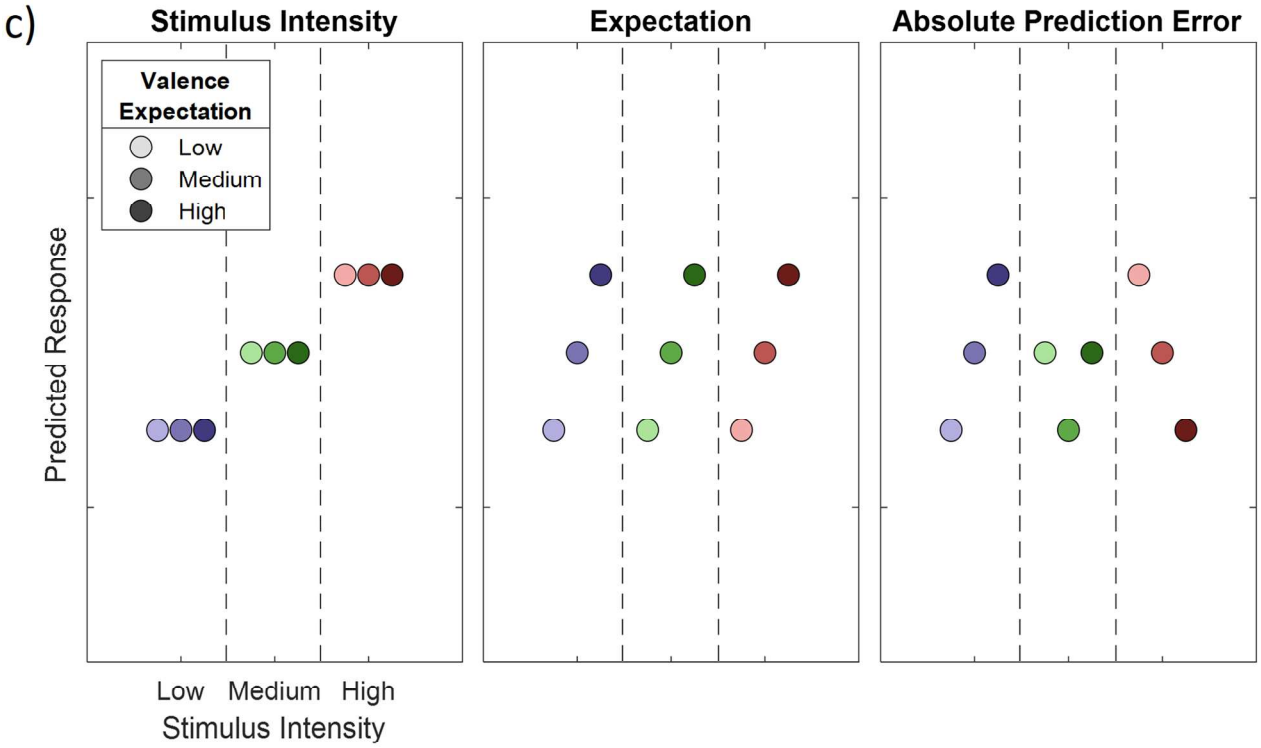



**Figure 1.** Overview of the study design. (**a**) Graphical representation of the trial structure. Each trial started with the presentation of a cue, indicating the stimulus intensity and modality of the following stimulus. After a jittered phase where only the fixation cross was shown, the stimulus (IAPS picture or pain) was presented. A rating phase (1–4) of the stimulus aversiveness followed. (**b**) Contingency table for all conditions for each cue-stimulus combination. Note that percentages are for all trials, therefore each row adds up to 1/6 (6 different cues). Orange fields indicate conditions included in the analysis, i.e. IAPS pictures where IAPS pictures were indicated by the color of the preceding cue. (**c**) Hypothetical response patterns based on Stimulus Intensity (INT; left), Expectation (EXP; middle) and Absolute Prediction Error (PE; right). The y-axis represents a hypothetical response variable (e.g. EEG power or rating). Each dot represents a different condition for each stimulus-cue combination. Blue colors represent low valence conditions, green colors represent medium valence conditions and red colors represent high valence conditions. Color intensities depict expectation level.

Prior to the actual EEG experiment, subjects participated in a behavioral training session. During this session, they were informed about the procedure and gave their written informed consent. The behavioral training session was implemented to avoid learning effects associated with the contingencies between the cues and the stimuli during the EEG session. Between two and three blocks were presented during the training session (without electrophysiological recordings). The experimenter assessed the performance after each block based on the percentage of successful catch trials and the ability to distinguish the three levels of aversiveness of each modality. The training session was terminated after the second block if participants were able to successfully label cues in 75% of the catch trials within the second block.

**EEG data acquisition.** EEG data were acquired using a 64-channel Ag/AgCl active electrode system (ActiCap64; BrainProducts) placed according to the extended 10–20 system[44]. Sixty electrodes were used of the most central scalp positions. The EEG was sampled at 500 Hz, referenced at FCz and grounded at Iz. For artifact removal, a horizontal, bipolar electrooculogram (EOG) was recorded using two of the remaining electrodes and placing them on the skin approximately 1 cm left from the left eye and right from the right eye at the height of the pupils. One vertical electrooculogram was recorded using one of the remaining electrodes centrally approximately 1 cm beneath the left eye lid and another electrode was fixated on the neck at the upper part of the left trapezius muscle to record an electromyogram (EMG).

**EEG preprocessing.** The parameters and procedures for the EEG preprocessing were adopted from the analysis of the pain sub-data set for reasons of comparability and consistency (see Strube et al. 2021, https://elifesciences.org/articles/62809 to view detailed comments from reviewers on these pre-processing steps)[43]. The data analysis was performed using the Fieldtrip toolbox for EEG/MEG-analysis[45]. EEG data were epoched and time-locked to the onset of the IAPS picture. Each epoch was centered (subtraction of the temporal mean) and detrended and included a time range of 3410 ms before and 2505 ms after trigger onset.

The data were band-pass filtered at 1–100 Hz, Butterworth, 4th order. EEG epochs were then visually inspected and trials contaminated by artifacts due to gross movements or technical artifacts were removed. Subsequently, trials contaminated by eye-blinks and movements were corrected using independent component analysis (ICA) with a single ICA per subject for all trials concatenated[46,47]. In all datasets, individual eye movements, showing a large EOG channel contribution and a frontal scalp distribution, were clearly seen in the removed independent components. Additionally, time–frequency decomposed ICA data were inspected at a single trial level for micro saccades and muscle artifacts, after z-transformation (only for artifact detection purposes) based on the mean and the standard deviation across all components separately for each frequency from 31 to 100 Hz. Time–Frequency representations were calculated using a sliding window multi-taper analysis with a window of 200 ms length, which was shifted over the data with a step size of 20 ms with a spectral smoothing of 15 Hz. Gamma artifact components were easily visible and were compared with the trial-by-trial time series representations of all ICA components. Specifically, single and separate muscle spikes and micro saccades were identified as columns or "clouds" in time–frequency plots. Using this procedure, up to 31 components were removed before remaining non-artefactual components were back-projected and resulted in corrected data. Subsequently, the data was re-referenced to a common average of all EEG channels and the previous reference channel FCz was re-used as a data channel.

Before time–frequency transformations for data analysis were performed on the cleaned data set, the time axis of single trials were shifted to create separate cue-locked and stimulus-locked datasets. Cue-locked data defines the onset of the cue as $t = 0$. Stimulus-locked data defines $t = 0$ as the onset of the picture stimulus. Frequencies up to 30 Hz (1 to 30 Hz in 1 Hz steps) were analyzed using a sliding Hanning-window Fourier transformation with a window length of 300 ms and a step-size of 50 ms. It should be noted that delta and theta frequencies are not ideally mapped with these tapers because of a short window length. For the analysis of frequencies higher than 30 Hz (31 to 100 Hz in 1 Hz steps) spectral analyses of the EEG data were performed using a sliding window multi-taper analysis. A window of 200 ms length was shifted over the data with a step size of 50 ms with a spectral smoothing of 15 Hz. Spectral estimates were averaged for each subject over trials. Afterwards, a z-baseline correction was performed based on a 500 ms baseline before cue onset to avoid differences in the baseline based on modulations of the signal by the anticipation period. For cue-locked data, a time frame ranging from −650 ms to −150 ms was chosen as a baseline. A distance from the cue onset to the baseline period of 150 ms was set because of the half-taper window length of 150 ms, i.e. data points between −150 ms and 0 ms are contaminated by the onset of the cue. For stimulus-locked trials, a variable cue duration (1500–1900 ms) was additionally taken into account, resulting in an according baseline from −2550 ms to −2050 ms from stimulus onset. For the baseline correction of time–frequency data, the mean and standard deviation were estimated for the baseline period (for each subject-channel-frequency combination, separately). The mean spectral estimate of the baseline was then subtracted from each data point, and the resulting baseline-centered values were divided by the baseline standard deviation[48].

**Predictive coding model.** Similar to a previous fMRI study[41] and our analysis of the pain subset of this dataset[43], our full model included three experimental within-subject factors (see Fig. 1c). The stimulus intensity factor (INT; see Fig. 1c; left column) models the measured response with a simple linear function of the stimulus intensity (−1, 0 and 1 for low, medium and high intensities, respectively). The expectation (EXP) factor was defined (see Fig. 1c; center column) linearly from the intensity predicted by the cue. Again, conditions with a low intensity cue were coded with a −1, conditions with a medium intensity cue with a 0 and conditions with a high intensity cue with a 1. The absolute prediction error factor (PE) resulted from the absolute difference of the expectation and actual stimulus intensity (see Fig. 1c; right column).

**Behavioral ratings.** Behavioral aversiveness ratings were averaged for all 3 × 3 cue-stimulus combinations over each participant, resulting in a 29 × 9 matrix (subject × condition). We tested for main effects across stimulus intensity, expectation, as well as prediction error using one-way repeated measures ANOVAs as implemented in MATLAB (see fitrm and ranova, Matlab version 2020a, The MathWorks). Post-hoc tests were performed on the repeated measures ANOVA models using Bonferroni corrections for multiple comparisons as implemented in MATLAB (see multcompare, Matlab version 2020a, The MathWorks).

**EEG data analysis.** The parameters and procedures for the EEG data analysis were adopted from the analysis of the pain sub-data set for reasons of comparability[43]. All statistical tests in electrode space were corrected for multiple comparisons using non-parametrical permutation tests of clusters[49].

We explored positive and negative time–frequency patterns associated with our variations of stimulus intensity, expectation and absolute prediction errors using one-way repeated measures ANOVAs as implemented in the Fieldtrip toolbox. A statistical value corresponding to $p = 0.05$ ($F(1,28) = 4.196$) obtained from the repeated measures ANOVA for each factor was used for clustering. Samples (exceeding the threshold of $F(1,28) = 4.196$) were clustered in connected sets on the basis of temporal (i.e. adjacent time points), spatial (i.e. neighboring electrodes) and spectral (i.e.± 1 Hz) adjacency. Further, clustering was restricted in a way that only samples were included in a cluster which had at least one significant neighbor in electrode space, i.e. at least one neighboring channel also had to exceed the threshold for a sample to be included in the cluster. Neighbors were defined by a template provided by the Fieldtrip toolbox corresponding to the used EEG montage.

Cluster tests were applied separately for low frequencies (1–30 Hz in 1 Hz steps) and high frequencies (31–100 Hz in 1 Hz steps) in a time frame from 0 (onset of visual stimulus) to 2000 ms (end of visual stimulus presentation) for stimulus-locked data and from 0 (onset of cue) to 1500 ms (visual stimulus onset) for cue-locked data. Stimulus-locked data was tested for stimulus intensity, expectation and absolute prediction errors factors. Cue-locked data was tested for the expectation factor.

Subsequently, a cluster value was defined as the sum of all statistical values of included samples. Monte Carlo sampling was used to generate 1000 random permutations of the design matrix and statistical tests were repeated in time–frequency-channel space with the random design matrix. The probability of a cluster from the original design matrix (p-value) was calculated by the proportion of random design matrices producing a cluster with a cluster value exceeding the original cluster where a p-value $< 0.05$ indicated a significant difference. This test was applied two-sided for negative and positive clusters. Positive and negative clusters were determined by the fixed factor estimate (average slope of all subjects) resulting from a simple linear regression analysis of the respective main effect, i.e. an average decrease with factor levels was coded negatively whereas an increase was coded positively.

Clusters of activity reaching statistical significance ($p < 0.05$) were further evaluated using post hoc tests, which were applied on the mean value of all time–frequency-channel combinations included in the cluster using Bonferroni corrections for multiple comparisons as implemented in MATLAB (see multcompare, Matlab version 2020a, The MathWorks).

## Results

**Behavioral data.** Participants experienced affective picture (or heat) stimuli which were probabilistically cued in terms of modality and intensity, evoking an expectation of modality and intensity. The subsequently applied stimuli were then rated on a visual analog scale (VAS) from 1–4. Our primary behavioral question was whether ratings are influenced by the experimental manipulation of stimulus intensity, expectation and absolute prediction errors.

To evaluate the main effects of stimulus intensity, expectation and absolute prediction error with regards to the valence of the IAPS pictures, we employed a repeated measures ANOVA of the behavioral data, which revealed significant effects for the main effect of stimulus intensity, i.e. the three levels of valence ($F(1,28) = 762.10$, $p < 0.001$). Post hoc analyses using the Bonferroni corrections for multiple comparisons for significance indicated that all three factor levels differed significantly, revealing higher ratings for high valence pictures ($M = 2.98$, $SD = 0.40$) vs medium valence pictures ($M = 1.70$, $SD = 0.30$) and medium valence pictures vs low valence pictures ($M = 1.09$, $SD = 0.07$; all $p < 0.001$).

The main effect for expectation on aversiveness ratings did not yield a significant effect ($F(1,28) = 1.46$, $p = 0.24$). However, the absolute difference between the cued intensity and the actual stimulus intensity (i.e. absolute prediction error), showed a significant effect on aversiveness ratings ($F(1,28) = 7.7$, $p = 0.01$). Post hoc tests indicated that the condition without PEs ($M = 1.95$, $SD = 0.24$) was significantly smaller than the high PE conditions ($M = 2.05$, $SD = 0.23$; $p < 0.001$). Also, the low PE condition ($M = 1.83$; $SD = 0.28$) was significantly smaller than the no PE condition ($p < 0.01$). In summary, aversiveness ratings were increasing with the degree of aversive valence of the presented picture stimuli. Moreover, these results demonstrate higher ratings when there was a mismatch between the degree of aversion signalized by the preceding cue and the actual stimulus content, i.e. high prediction errors are related to higher aversiveness ratings. The results regarding picture stimuli are summarized in Table 1. See Fig. 2 for a descriptive rain cloud plot of behavioral ratings for each condition, main effects plot for each factor and single subject parameter estimates, showing significant positive intensity and prediction error factors.

**EEG Intensity.** EEG analysis were performed in the same way as the pain sub-data set[43]. We tested our EEG time–frequency data for a main effect of the valence of the aversive IAPS pictures in the context of a correctly cued modality (i.e. an IAPS picture was expected and received). In order to do so, we performed a repeated measures ANOVA on the time–frequency representation of the EEG data on low frequencies (1–30 Hz) and

| | Stimulus intensity (INT) | | Cued intensity (EXP) | | Absolute prediction error (PE) | |
|---|---|---|---|---|---|---|
| Factor | *F(1,28)* | *p* | *F(1,28)* | *p* | *F(1,28)* | *p* |
| Behaviroal ratings | 762.10 | < .001 | 1.46 | .24 | 7.7 | .01 |

**Table 1.** Main effects of stimulus intensity, expectation and absolute prediction errors on subjective aversiveness ratings in affective picture conditions.

high frequencies (31–100 Hz) separately using a cluster correction criterion to address the multiple comparisons problem (see "Methods" for details). Any significant cluster—composed of neighboring data points in time, frequency and space—would indicate a neuronal oscillatory representation of variations in stimulus intensity in a given frequency band.

In the low frequency (1–30 Hz) range, we observed one significant negative cluster of activity ($p < 0.001$) indicating a negative association of IAPS valence and power in the alpha-to-beta range (See Fig. 3 for a time–frequency representation, a main effects plot and single subject parameter estimates of the INT cluster). Specifically, this negative cluster included samples in a time range from 0 to 2000 ms after IAPS stimulus onset in a frequency range from 1 to 30 Hz. All channels included samples of the negative low frequency stimulus intensity cluster. Bonferroni corrected post hoc tests applied on the mean value of all time–frequency-channel combinations included in the INT cluster revealed that all comparisons, i.e. low valence ($M = -0.38$, $SD = 0.80$) vs medium valence ($M = -0.85$, $SD = 0.80$), medium valence vs high valence ($M = -1.08$, $SD = 0.82$) and low valence vs high valence were significant (all $p < 0.05$), i.e. higher picture valence was related to lower alpha-to-beta power.

In conclusion, these results indicate that a higher picture valence is associated with decreased alpha-to-beta band power (see Fig. 4 for a rain cloud plot of average EEG power at the INT cluster). No effect was observed for higher frequencies between 31 and 100 Hz.

**Expectation.** In a next step, we investigated the representation of the expectation factor (EXP) in our repeated-measures model, again for low frequencies (1–30 Hz) and high frequencies (31–100 Hz) separately in the IAPS stimulus-locked and cue-locked time–frequency representation of the EEG data.

This analysis revealed one significant negative cluster in the low frequency range (1–30 Hz) after IAPS stimulus onset, indicating a negative association of cued intensity (EXP) and power in this frequency range ($p < 0.05$). The expectation cluster ($p = 0.017$) included samples from time points ranging from 550 to 1750 ms after IAPS stimulus onset and included frequencies from 3 to 30 Hz. All channels included samples of the negative low frequency EXP cluster (See Fig. 5 for a time–frequency representation, a main effects plot and single subject parameter estimates of the EXP cluster). Post hoc tests revealed that all comparisons, i.e. low valence expectation ($M = -0.77$, $SD = 0.65$) vs medium valence expectation ($M = -1.11$, $SD = 0.68$), medium valence expectation vs high valence expectation ($M = -1.33$, $SD = 0.62$) and low valence expectation vs high valence expectation were significant (all $p < 0.001$, Bonferroni-corrected), showing higher valence expectation was related to lower alpha-to-beta power.

A cluster analysis of the expectation factor in cue-locked EEG data (from 1 to 30 Hz for low frequencies and 31–100 Hz for gamma frequencies; from 0 to 1500 ms), did not reveal any significant cluster of activity associated with changes in EXP (all $p > 0.05$,). See Supplementary Fig. 1 for time–frequency representations for low, medium and high valence expectation conditions.

In conclusion, these results indicate that a higher valence expectation is associated with decreased alpha-to-beta band power during stimulus presentation.

**Absolute prediction errors.** Finally, we investigated the representation of absolute prediction errors (PE) in our repeated-measures model for low frequencies (1–30 Hz) and high frequencies (31–100 Hz) separately in the IAPS stimulus-locked time–frequency representation of the EEG data. This analysis revealed two significant adjacent positive cluster after IAPS stimulus onset, indicating a positive modulation of EEG power by absolute prediction errors (PE) ($p < 0.05$).

One positive prediction error cluster was found in the low frequency range (1–30 Hz) ($p < 0.001$) and included samples from time points ranging from 0 to 2000 ms after IAPS stimulus onset and included frequencies from 1 to 30 Hz. All channels included samples of the low frequency absolute prediction error cluster (see Fig. 6 for a time–frequency representation, a main effects plot and single subject parameter estimates of the low frequency PE cluster). Here, post hoc tests revealed that conditions without prediction errors ($M = -1.55$, $SD = 0.82$) were associated with significantly lower alpha-to-beta power than both, low ($M = -0.80$, $SD = 0.55$) and high PE ($M = -0.68$, $SD = 0.54$) conditions (all $p < 0.001$) whereas medium and high PE conditions did not differ in alpha-to-beta power ($p = 0.1$).

In the high frequency range (31–100 Hz) representing gamma activity one positive prediction error cluster was observed ($p < 0.001$) and included samples ranging from 0 to 2000 ms after IAPS stimulus onset and from 31 to 73 Hz. All channels included samples of the high frequency absolute prediction error cluster (See Fig. 7 for a time–frequency representation, a main effects plot and single subject parameter estimates of the INT cluster). Post hoc tests revealed that all comparisons were significant (all $p < 0.01$, Bonferroni-corrected) and conditions without PEs ($M = -0.84$, $SD = 0.46$) were associated with a significantly lower gamma power than low PE conditions ($M = -0.36$, $SD = 0.31$), and low PE conditions were associated with a lower gamma power than high PE conditions ($M = -0.19$, $SD = 0.34$).

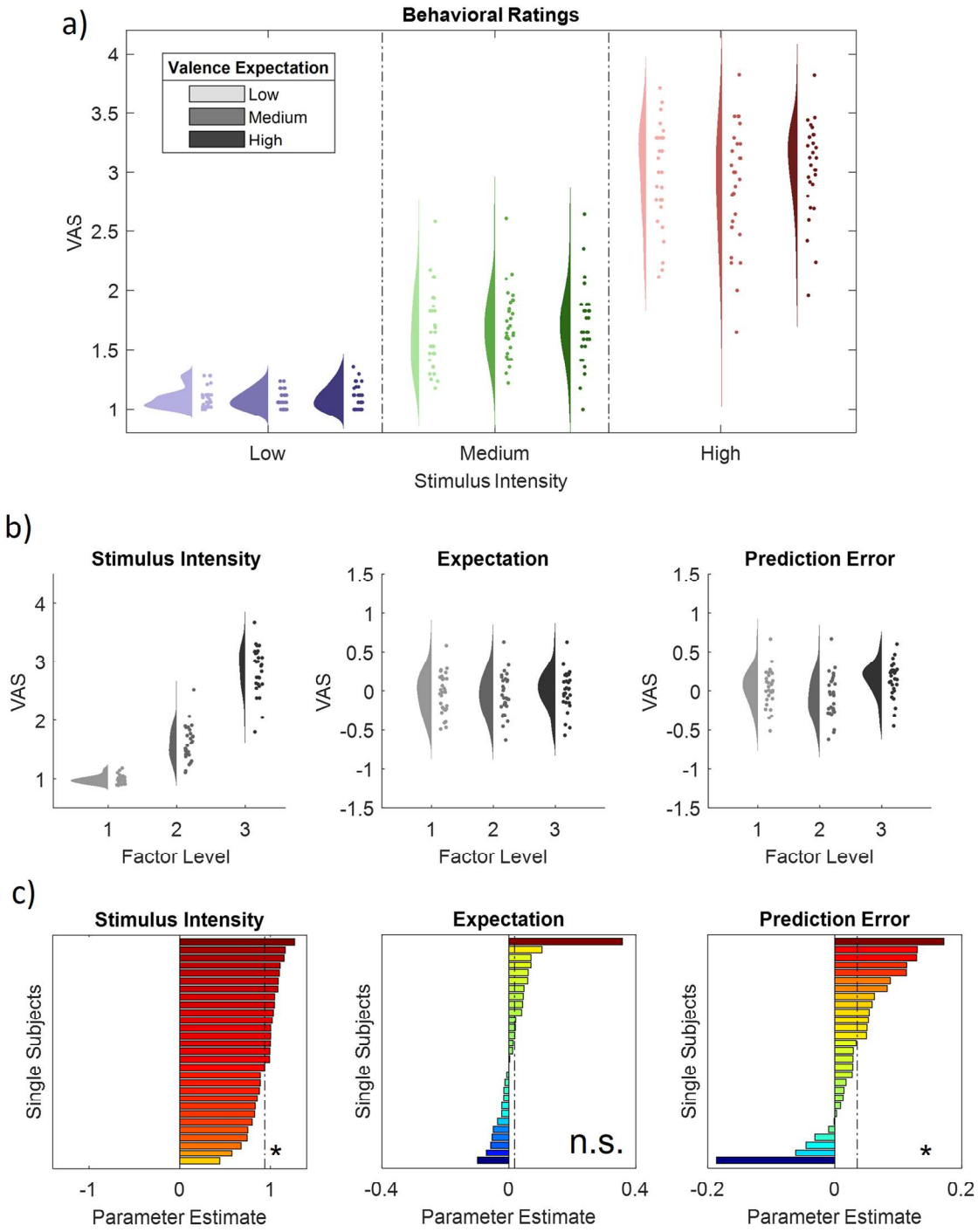



**Figure 2.** Ratings for IAPS picture stimuli. (**a**) Raincloud plots representing single subject ratings for all 9 congruent conditions (expect a picture and receive a picture). VAS (Visual Analog Scale) represents the rating on a 1–4 rating scale. Blue colors represent low valence IAPS picture stimuli, green colors medium valence IAPS picture stimuli and red colors high valence IAPS picture stimuli. The data show both an effect of stimulus intensity (increase from blue to green to red), but also a significant positive effect of absolute prediction errors. (**b**) Main effect plots for the stimulus intensity ($F(1,28) = 762.10$, $p < .01$), expectation ($F1,28) = 1.46$, $p = .24$) and prediction error ($F1,28) = 7.7$, $p < .05$) factors (from left to right) showing single subject values and distributions on the response, partialling out (for display purposes) the effects of the other predictors (e.g. EXP and PE were partialled out for the main effect plot of INT) for all three factor levels (increasing from left to right). (**c**) Bars represent the estimated slope for each subject and factor (stimulus intensity, expectation and prediction error from left to right). The dashed line represents the fixed factor estimate (average slope of all subjects). Hot colors represent a positive slope (increases with factor levels) and cold colors a negative slope (decreases with factor levels).

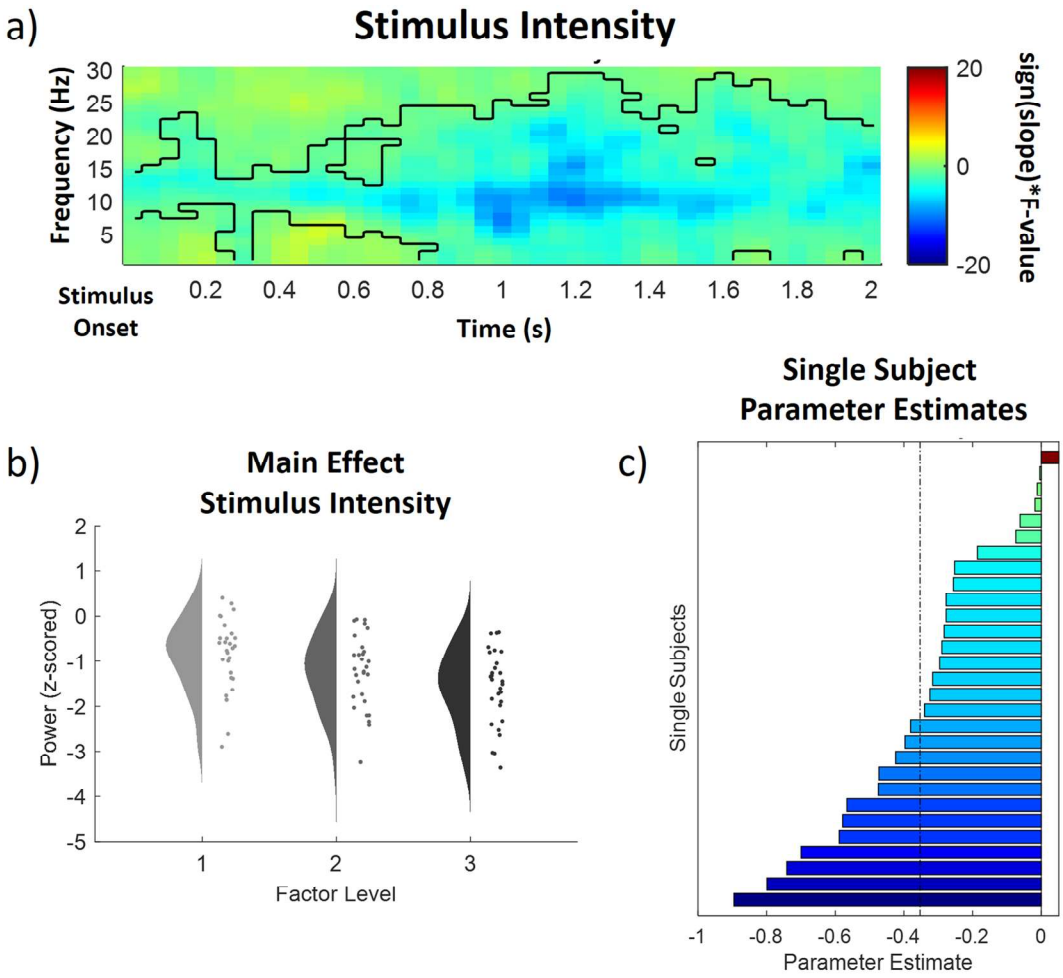



**Figure 3.** Time–frequency representation (**a**), main effect plot (**b**) and single subject parameter estimates (**c**) for the significant stimulus intensity (INT) cluster ($p < .001$, cluster-corrected), showing a decrease of alpha-to-beta power with an increased aversiveness of the stimulus. Time–frequency representations (**a**) are composed of the statistical F-values of the repeated measures ANOVA averaged over all channels. The significant cluster is outlined. Hot colors represent a positive slope (increases with factor levels) and cold colors a negative slope (decreases with factor levels). The main effect plot (**b**) for the INT cluster summarizes single subject values and distributions on the response, partialling out the effects of the respective other predictors (i.e. EXP and PE were averaged out for the main effect plot of INT) for all three factor levels (increasing from left to right). Single subject parameter estimates (**c**) are based on a linear regression of single subject values of each factor level. The dashed line represents the fixed factor estimate (average slope of all subjects).

In summary, these results suggest an increase in alpha-to-beta and low gamma band power to be associated with expectation violations (i.e. absolute prediction errors), resulting from a mismatch of the cued intensity and the actual valence of the IAPS stimulus. Even though the parameters of our cluster analysis resulted in two separate clusters of activity, these clusters are connected in time, frequency and space which suggests this activity to be related to one single cluster.

## Discussion

Using a probabilistic cue paradigm with affective pictures of different valence levels, our data showed a clear discriminability of valence based on behavioral ratings and EEG time frequency patterns. Valence ratings were positively modulated by high prediction errors, supporting the hypothesis that prediction errors are linked to higher (negative) valence[40]. With regards to the EEG data, we observed one cluster of activity to be negatively correlated with the valence of the IAPS material in the alpha-to-beta band. Most importantly, our analysis also revealed expectations and violations of expectations (i.e. prediction errors) to be involved in the alpha-to-beta ERD and gamma power modulations.

Firstly, we hypothesized that alpha-to-beta ERD responses should be modulated by expectations. Additionally, we expected a modulation of these frequencies during the anticipation period from cue onset to the onset of the IAPS stimulus. Here, we found higher alpha-to-beta ERD associated with higher valence expectations during stimulus presentation, whereas we found no differences during the anticipation period.

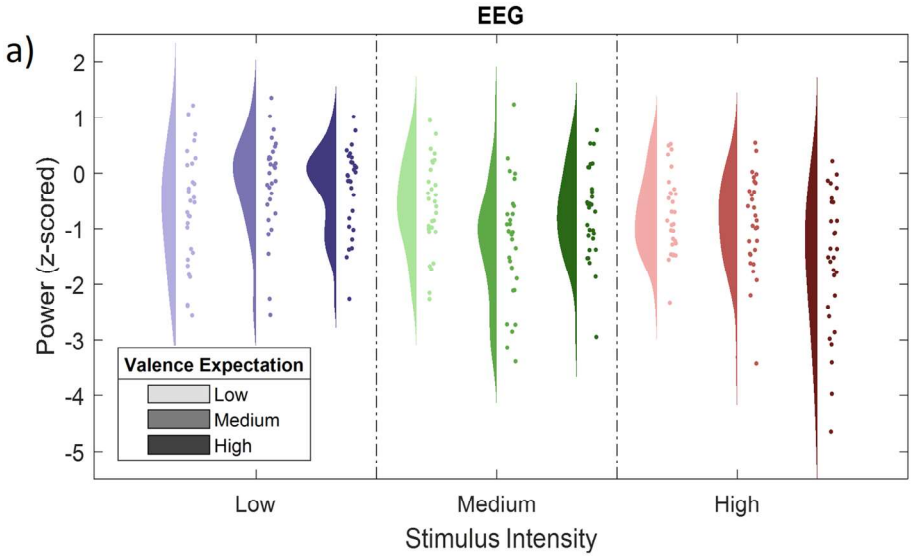



**Figure 4.** EEG activity at the significant INT cluster. (**a**) Scatter plots representing single subject EEG power (averaged over all samples included in the significant INT cluster) for all 9 congruent conditions (expect a picture and receive a picture) and according probability distributions averaged over all significant samples included in the negative INT cluster (0–2000 ms; 1-30 Hz).

Secondly, we hypothesized that surprise should lead to an increase of gamma ERS when there was a mismatch between the anticipated degree of aversion and the actual aversive quality of the picture. In contrast, if the negative valence or aversive quality is contributing to the gamma ERS effect, we expected an increase of gamma ERS with higher aversion regardless of the anticipated degree of aversion. Here, we provide evidence for gamma activity related to surprise as higher gamma power was associated with absolute prediction errors, whereas higher picture valence did not manifest in gamma power increases.

Our findings are in agreement with reports of decreases of alpha band power with unpleasant images and emotional arousal[19–28]. Even though many studies observed a decrease in power in the alpha- and lower beta band, some studies observed an increase with increased valence[28,30,50–52]. Interestingly, we found alpha-to-beta increases in power to be related to expectation violations as well as alpha-to-beta decreases associated with expected valence during the presentation of the IAPS stimulus. Anticipation of negative pictures enhances neural responses to the pictures[53,54] during encoding of the emotional content, which is well in line with our findings of increased ERD with higher valence expectations.

Conversely, anticipation of aversive images did not manifest as differences after the presentation of the cue. It has been shown that in the anticipation period for affective images, alpha ERD preceding an anticipated negative image was larger as compared to a positive image[34]. Also, negative anticipation of affective images have been associated with the activation of the right prefrontal cortex in fMRI studies[55,56]. Interestingly, activation of brain areas associated with negative anticipation is decreased when anticipation of negative emotion is uncertain[57]. Here, all cues were to a large degree uncertain (after all, only 60% of all cues predicted the intensity correctly), which could explain that we could not detect expectation signals based on uncertainty of the anticipation. Alpha-to-beta band activity has been specifically implicated in the processing of top-down expectation signals[17,18]. Beta activity has also been linked to top-down prediction signals in the visual perception of causal events[58]. Here, we find alpha-to-beta activity associated with expectation signal only during stimulus presentation, suggesting that a representation of the prediction is reinstantiated during stimulus presentation.

EEG desynchronization is considered a reliable correlate of excited neural structures or activated cortical areas, while synchronization within the alpha band is hypothesized to be an electrophysiological correlate of deactivated cortical areas[59] (see Pfurtscheller et al., 1996 for a review). An alternative view suggests increased alpha activity to be associated with active inhibition rather than passive inactivity[60–65]. More specifically, it has been suggested that alpha activity represents an attentional suppression mechanism when objects or features need to be specifically ignored or selected against[60]. Moreover, event related alpha synchronization is obtained over sites that probably exert top-down control and hence it has been assumed that alpha synchronization reflects a top-down process of inhibitory control[63].

In this sense, inhibition is a mechanism for gating the flow of information throughout the brain which is mediated by alpha activity[61,62,65]. In our study, two effects come to play in the alpha-to-beta band, which are relevant with regard to this hypothesis: Firstly, alpha band activity shows a negative relationship with expected stimulus intensity, suggesting less inhibition (i.e. more attention to this information) of highly aversive (potentially negative or threatening) visual stimulation. Secondly, prediction errors resulted in increased alpha band power, i.e. a positive relationship. In this sense, incongruent trials would be attentionally suppressed and the features would be specifically ignored or selected against. This is because our alpha-to-beta prediction error follows a pattern of higher alpha-to-beta power with higher prediction errors. In this paradigm, the probabilistic characteristics of

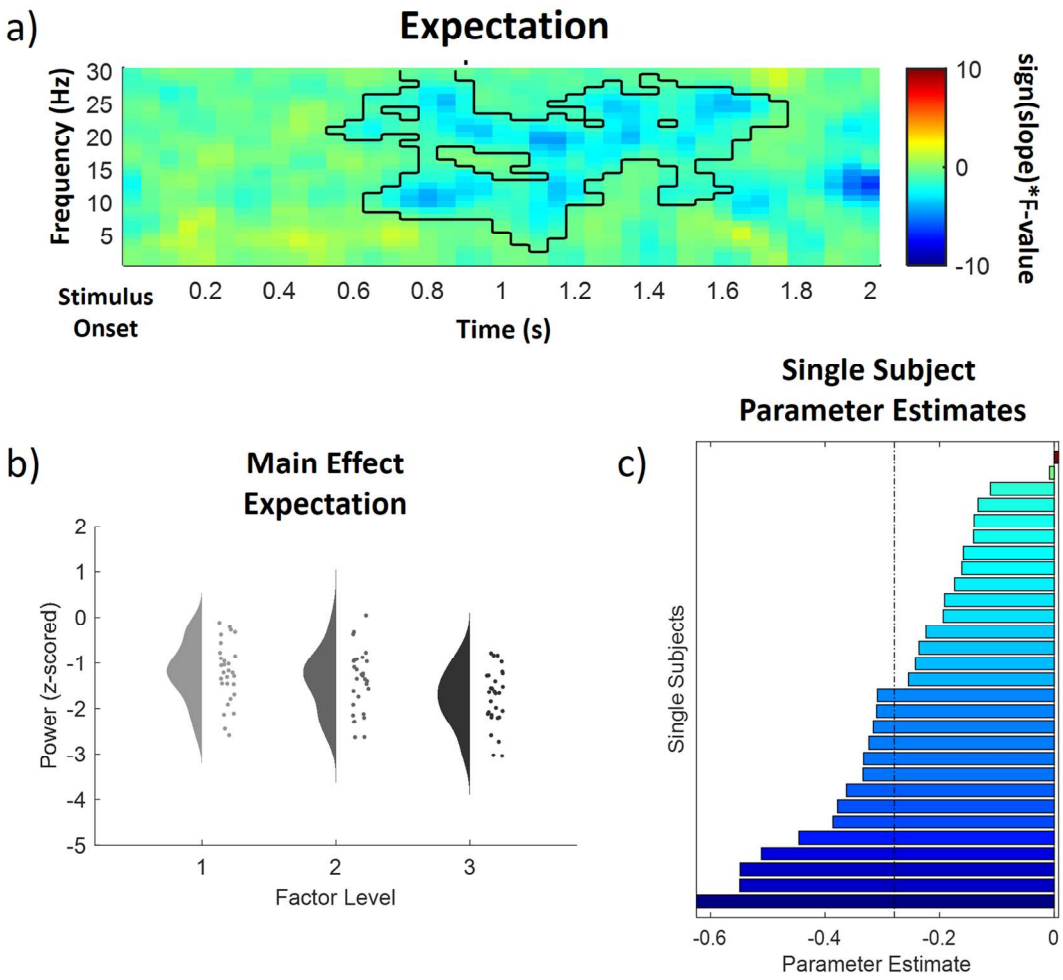



**Figure 5.** Time–frequency representation (**a**), main effect plot (**b**) and single subject parameter estimates (**c**) for the significant expectation (EXP) cluster ($p < .05$, cluster-corrected), showing a decrease of alpha-to-beta power with an increased expected valence of the stimulus. Time–frequency representations (**a**) are composed of the statistical F-values of the repeated measures ANOVA averaged over all channels. The significant cluster is outlined. Hot colors represent a positive slope (increases with factor levels) and cold colors a negative slope (decreases with factor levels). The main effect plot (**b**) for the EXP cluster summarizes single subject values and distributions on the response, partialling out the effects of the respective other predictors (i.e. INT and PE were averaged out for the main effect plot of EXP) for all three factor levels (increasing from left to right). Single subject parameter estimates (**c**) are based on a linear regression of single subject values of each factor level. The dashed line represents the fixed factor estimate (average slope of all subjects).

the cue did not change during the experiment and were learned before EEG measurements. It has been shown that the update of predictions is associated with beta ERD[66]. Here, an update of predictions based on unlikely events would reflect a change in predictions, even though the actual probabilities did not change. Following this thought, the update of predictions might be suppressed which manifests as higher alpha-to-beta power in prediction error conditions. This is in line with a proposed role of beta activity in actively maintaining the current cognitive set or the status quo[67].

Beta oscillations have also been suggested to be associated with temporal reactivation of neural representations[68]. Beta modulations have been shown in working memory tasks, in which past information is brought into the focus of attention[68–70]. Here, conditions with prediction errors might be related to a similar process, where a mismatch was evaluated by a focus on the information of the cue stored in working memory. Our manipulation of prediction errors was associated with an increase in alpha-to-beta power: this suggests top-down processes (working memory and suppression) instead of bottom-up processes to be at play at alpha-to-beta frequencies associated with our prediction error factor.

In predictive coding, gamma activity has been specifically associated with prediction error responses[17,18] and has been associated with bottom-up prediction errors in the visual processing of causal events[58]. Here, we found two clusters associated with prediction errors, firstly in the alpha to beta range and secondly in the gamma range. The gamma cluster needs to be interpreted with caution, as it might be affected by spectral smearing from the alpha-to-beta cluster.

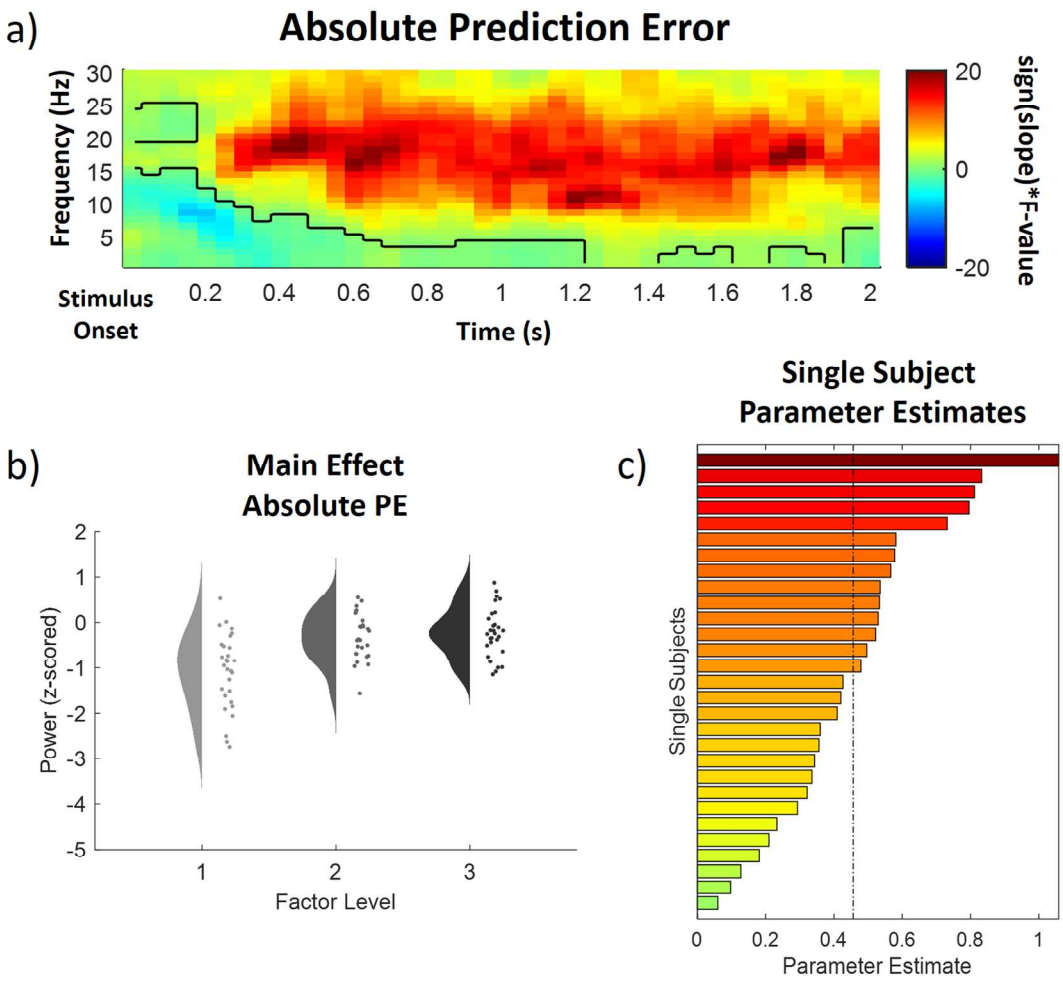



**Figure 6.** Time–frequency representation (**a**), main effect plot (**b**) and single subject parameter estimates (**c**) for the significant low frequency absolute prediction error (PE) cluster ($p < .001$, cluster-corrected), showing an increase of alpha-to-beta power with prediction errors. Time–frequency representations (**a**) are composed of the statistical F-values of the repeated measures ANOVA averaged over all channels. The significant cluster is outlined. Hot colors represent a positive slope (increases with factor levels) and cold colors a negative slope (decreases with factor levels). The main effect plot (**b**) for the PE cluster summarizes single subject values and distributions on the response, partialling out the effects of the respective other predictors (i.e. INT and EXP were averaged out for the main effect plot of PE) for all three factor levels (increasing from left to right). Single subject parameter estimates (**c**) are based on a linear regression of single subject values of each factor level. The dashed line represents the fixed factor estimate (average slope of all subjects).

In predictive coding, an improved causal model by learning improves top-down predictions which consequently lead to a reduction of bottom-up prediction error signals[71]. If we interpret both PE clusters (in the alpha- to beta range and in the gamma range) as incorporating different processes which are encoded simultaneously at different frequencies, our gamma PE cluster might be a manifestation of bottom-up prediction error signals. In predictive coding, gamma activity depends on the match between expectations and bottom-up input[17] and is in this sense an assessment of sensory predictions[72]. In this study, we could directly assess the difference between expectations and bottom-up sensory input, resulting in differences in the gamma range. In summary, this would imply that top-down working memory demands and the suppression of prediction updates were encoded in the alpha-to-beta range whereas bottom-up prediction error signals were simultaneously encoded in the gamma range.

In the formulation of predictive coding, an important function of emotional valence turns out to regulate the learning rate of the causes of sensory inputs. Specifically it has been proposed that a violation of expectation leads to a (qualitatively) negative valence and an increase of the learning rate, while fulfilled expectations are associated with positive valence and a decrease of the learning rate[40]. Absolute prediction errors are also integral part of formal learning models. In the Pearce Hall model[73], the absolute error promotes changes in associative strength (i.e. learning rate) such that large absolute prediction errors (surprises) prompt the model to rapidly adapt by increasing its learning rate. If emotions can be derived from a predictive coding function, visual processing of affective pictures can be seen as a simplified model of predictive coding processes in emotion.

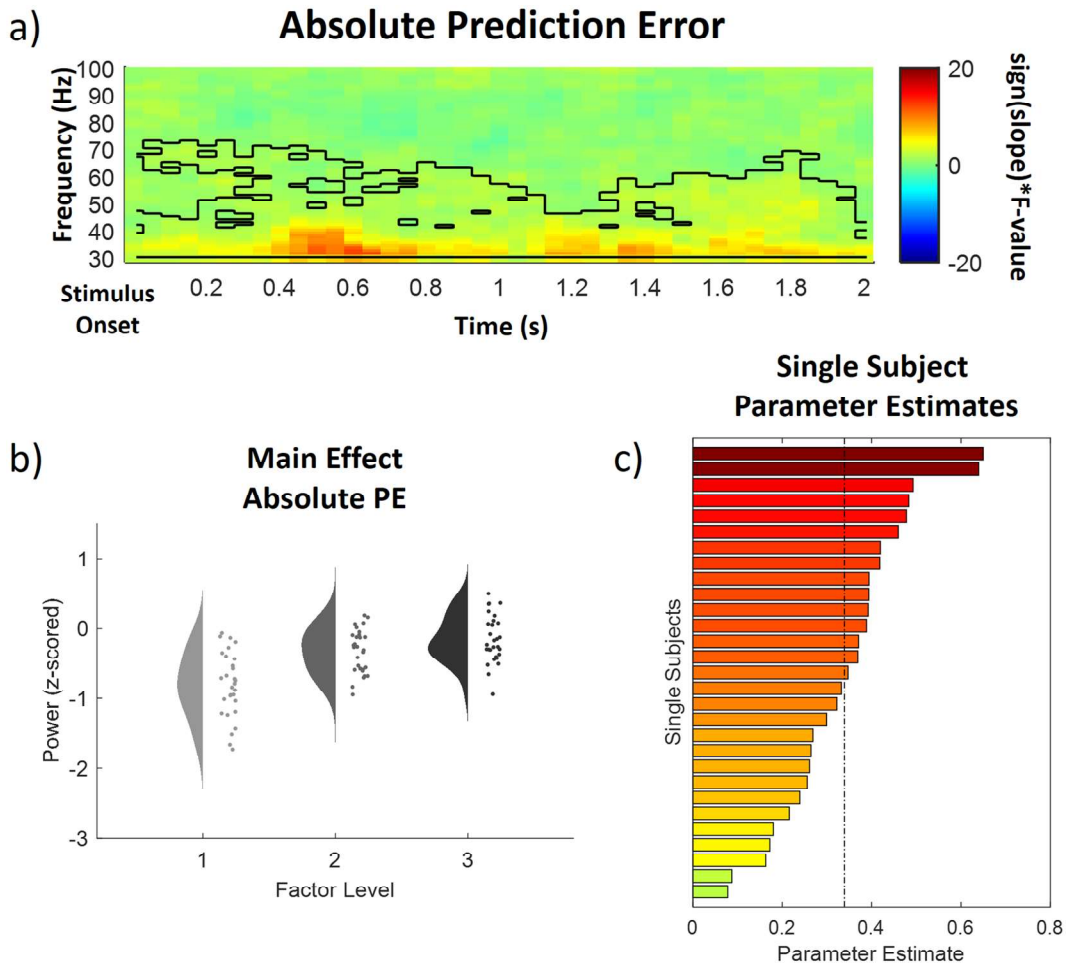



**Figure 7.** Time–frequency representation (**a**), main effect plot (**b**) and single subject parameter estimates (**c**) for the significant high frequency absolute prediction error (PE) cluster ($p < .001$, cluster-corrected), showing an increase of gamma power with prediction errors. Time–frequency representations (**a**) are composed of the statistical F-values of the repeated measures ANOVA averaged over all channels. The significant cluster is outlined. Hot colors represent a positive slope (increases with factor levels) and cold colors a negative slope (decreases with factor levels). The main effect plot (**b**) for the PE cluster summarizes single subject values and distributions on the response, partialling out the effects of the respective other predictors (i.e. INT and EXP were averaged out for the main effect plot of PE) for all three factor levels (increasing from left to right). Single subject parameter estimates (**c**) are based on a linear regression of single subject values of each factor level. The dashed line represents the fixed factor estimate (average slope of all subjects).

## Limitations

This study has been designed in close analogy to a previous fMRI study to unravel the temporal dynamics of expectation and prediction errors and we decided to use the same experimental paradigm[41]. We therefore decided to also keep the sample characteristics similar and restricted the sample to male participants, which means that we cannot generalize our results to the population. Future studies should investigate samples including female participants. This would also allow to investigate sex effects with respect to expectation and prediction error effects in affective picture processing. The restriction to negative valence stimuli in this study limit the generalizability of our findings. Future studies could explicitly investigate positively valenced stimuli in the context of predictive coding.

## Summary

Our data show that key variables required for affective picture processing in the context of a generative model (i.e. predictive coding) are correlated with event-related alpha-to-beta and gamma activity. Alpha-to-beta activity was (negatively) modulated by valence expectations and stimulus valence, whereas prediction errors (positively) modulated responses from alpha-to-gamma frequencies. Alpha-to-beta increases associated with the mismatch of stimulus valence and expected valence imply working memory demands as well as the suppression of prediction updates, whereas gamma increases suggest a role of bottom-up processing of prediction errors.

Received: 24 July 2021; Accepted: 22 November 2021
Published online: 06 December 2021

## References


1. Anaki, D., Zion-Golumbic, E. & Bentin, S. Electrophysiological neural mechanisms for detection, configural analysis and recognition of faces. *NeuroImage* **37**, 1407–1416 (2007).
2. Zion-Golumbic, E., Kutas, M. & Bentin, S. Neural dynamics associated with semantic and episodic memory for faces: Evidence from multiple frequency bands. *J. Cogn. Neurosci.* **22**, 263–277 (2010).
3. Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* **36**, 181–204 (2013).
4. Huang, Y. & Rao, R. P. N. Predictive coding. *Wiley Interdiscip. Rev. Cogn. Sci.* **2**, 580–593 (2011).
5. Knill, D. C. & Pouget, A. The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends Neurosci.* **27**, 712–719 (2004).
6. Friston, K. The free-energy principle: A unified brain theory?. *Nat. Rev. Neurosci.* **11**, 127–138 (2010).
7. Balconi, M. & Lucchiari, C. Consciousness and arousal effects on emotional face processing as revealed by brain oscillations. A gamma band analysis. *Int. J. Psychophysiol.* **67**, 41–46 (2008).
8. Jung, J. *et al.* Intracerebral gamma modulations reveal interaction between emotional processing and action outcome evaluation in the human orbitofrontal cortex. *Int. J. Psychophysiol.* **9**, 2 (2011).
9. Keil, A. *et al.* Effects of emotional arousal in the cerebral hemispheres: A study of oscillatory brain activity and event-related potentials. *Clin. Neurophysiol.* **112**, 2057–2068 (2001).
10. Keil, A., Stolarova, M., Moratti, S. & Ray, W. J. Adaptation in human visual cortex as a mechanism for rapid discrimination of aversive stimuli. *Neuroimage* **36**, 472–479 (2007).
11. Luo, Q., Holroyd, T., Jones, M., Hendler, T. & Blair, J. Neural dynamics for facial threat processing as revealed by gamma band synchronization using MEG. *Neuroimage* **34**, 839–847 (2007).
12. Martini, N. *et al.* The dynamics of EEG gamma responses to unpleasant visual stimuli: From local activity to functional connectivity. *Neuroimage* **60**, 922–932 (2012).
13. Matsumoto, A., Ichikawa, Y., Kanayama, N., Ohira, H. & Iidaka, T. Gamma band activity and its synchronization reflect the dysfunctional emotional processing in alexithymic persons. *Psychophysiology* **43**, 533–540 (2006).
14. Oya, H., Kawasaki, H., Howard, M. A. & Adolphs, R. Electrophysiological responses in the human amygdala discriminate emotion categories of complex visual stimuli. *J. Neurosci.* **22**, 9502–9512 (2002).
15. Sato, W. *et al.* Rapid amygdala gamma oscillations in response to fearful facial expressions. *Neuropsychologia* **49**, 612–617 (2011).
16. Yoshino, A. *et al.* Sadness enhances the experience of pain and affects pain-evoked cortical activities: An MEG study. *J. Pain* **13**, 628–635 (2012).
17. Arnal, L. H. & Giraud, A.-L. Cortical oscillations and sensory predictions. *Trends Cogn. Sci.* **16**, 390–398 (2012).
18. Bastos, A. M. *et al.* Canonical microcircuits for predictive coding. *Neuron* **76**, 695–711 (2012).
19. Cesarei, A. D. & Codispoti, M. Affective modulation of the LPP and α-ERD during picture viewing. *Psychophysiology* **48**, 1397–1404 (2011).
20. Cui, Y. *et al.* Alpha oscillations in response to affective and cigarette-related stimuli in smokers. *Nicotine Tob. Res.* **15**, 917–924 (2013).
21. Ferrari, V., Bradley, M. M., Codispoti, M. & Lang, P. J. Massed and distributed repetition of natural scenes: Brain potentials and oscillatory activity. *Psychophysiology* **52**, 865–872 (2015).
22. Lee, J. Y., Lindquist, K. A. & Nam, C. S. Emotional granularity effects on event-related brain potentials during affective picture processing. *Front. Hum. Neurosci.* **11**, 2 (2017).
23. Meng, X. *et al.* EEG oscillation evidences of enhanced susceptibility to emotional stimuli during adolescence. *Front. Psychol* **7**, 2 (2016).
24. Messerotti Benvenuti, S., Buodo, G., Mennella, R., Dal Bò, E. & Palomba, D. Appetitive and aversive motivation in depression: The temporal dynamics of task-elicited asymmetries in alpha oscillations. *Sci. Rep.* **9**, 17129 (2019).
25. Schneider, D., Göddertz, A., Haase, H., Hickey, C. & Wascher, E. Hemispheric asymmetries in EEG alpha oscillations indicate active inhibition during attentional orienting within working memory. *Behav. Brain Res.* **359**, 38–46 (2019).
26. Schneider, T. R. *et al.* Modulation of neuronal oscillatory activity in the beta- and gamma-band is associated with current individual anxiety levels. *NeuroImage* **178**, 423–434 (2018).
27. Schubring, D. & Schupp, H. T. Affective picture processing: Alpha- and lower beta-band desynchronization reflects emotional arousal. *Psychophysiology* **56**, e13386 (2019).
28. Schubring, D. & Schupp, H. T. Emotion and brain oscillations : High arousal is associated with decreases in alpha- and lower beta-band power. *Cereb. Cortex* https://doi.org/10.1093/cercor/bhaa312 (2020).
29. Boucher, O. *et al.* Spatiotemporal dynamics of affective picture processing revealed by intracranial high-gamma modulations. *Hum. Brain Mapp.* **36**, 16–28 (2015).
30. Güntekin, B. & Tülay, E. Event related beta and gamma oscillatory responses during perception of affective pictures. *Brain Res.* **1577**, 45–56 (2014).
31. Müller, M. M., Gruber, T. & Keil, A. Modulation of induced gamma band activity in the human EEG by attention and visual information processing. *Int. J. Psychophysiol.* **38**, 283–299 (2000).
32. Müller, M. M., Keil, A., Gruber, T. & Elbert, T. Processing of affective pictures modulates right-hemispheric gamma band EEG activity. *Clin. Neurophysiol.* **110**, 1913–1920 (1999).
33. Yang, K. *et al.* High gamma band EEG closely related to emotion: Evidence from functional network. *Front. Hum. Neurosci.* **14**, 2 (2020).
34. Onoda, K. *et al.* Anticipation of affective images and event-related desynchronization (ERD) of alpha activity: An MEG study. *Brain Res.* **1151**, 134–141 (2007).
35. Kube, T., Schwarting, R., Rozenkrantz, L., Glombiewski, J. A. & Rief, W. Distorted cognitive processes in major depression: A predictive processing perspective. *Biol. Psychiat.* **87**, 388–398 (2020).
36. Chekroud, A. M. Unifying treatments for depression: An application of the free energy principle. *Front. Psychol.* **6**, 2 (2015).
37. Barrett, L. F., Quigley, K. S. & Hamilton, P. An active inference theory of allostasis and interoception in depression. *Phil. Trans. R. Soc. B* **371**, 20160011 (2016).
38. Clark, J. E., Watson, S. & Friston, K. J. What is mood? A computational perspective. *Psychol. Med.* **48**, 2277–2284 (2018).
39. Smith, R. *et al.* A Bayesian computational model reveals a failure to adapt interoceptive precision estimates across depression, anxiety, eating, and substance use disorders. *medRxiv* https://doi.org/10.1101/2020.06.03.20121343 (2020).
40. Joffily, M. & Coricelli, G. Emotional valence and the free-energy principle. *PLoS Comput. Biol.* **9**, e1003094 (2013).
41. Fazeli, S. & Büchel, C. Pain-related expectation and prediction error signals in the anterior insula are not related to aversiveness. *J. Neurosci.* **38**, 6461–6474 (2018).
42. Bradley, M. M. & Lang, P. J. The international affective picture system (IAPS) in the study of emotion and attention. In *Handbook of Emotion Elicitation and Assessment* 29–46 (Oxford University Press, 2007).

43. Strube, A., Rose, M., Fazeli, S. & Büchel, C. The temporal and spectral characteristics of expectations and prediction errors in pain and thermoception. *eLife* **10**, e62809 (2021).
44. Klem, G. H., Lüders, H. O., Jasper, H. H. & Elger, C. The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. *Electroencephalogr. Clin. Neurophysiol Suppl* **52**, 3–6 (1999).
45. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, 156869 (2011).
46. Makeig, S., Bell, A., Jung, T.-P. & Sejnowski, T. Independent component analysis of electroencephalographic data. *Adv. Neural Inform. Sci.* **8**, 2 (1996).
47. Jung, T. P. *et al.* Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* **37**, 163–178 (2000).
48. Grandchamp, R. & Delorme, A. Single-trial normalization for event-related spectral decomposition reduces sensitivity to noisy trials. *Front. Psychol.* **2**, 2 (2011).
49. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).
50. Aftanas, L. I., Varlamov, A. A., Pavlov, S. V., Makhnev, V. P. & Reva, N. V. Time-dependent cortical asymmetries induced by emotional arousal: EEG analysis of event-related synchronization and desynchronization in individually defined frequency bands. *Int. J. Psychophysiol.* **44**, 67–82 (2002).
51. Mennella, R. *et al.* The two faces of avoidance: Time-frequency correlates of motivational disposition in blood phobia. *Psychophysiology* **54**, 1606–1620 (2017).
52. Popov, T., Miller, G. A., Rockstroh, B. & Weisz, N. Modulation of α power and functional connectivity during facial affect recognition. *J. Neurosci.* **33**, 6018–6026 (2013).
53. Lin, H., Xiang, J., Li, S., Liang, J. & Jin, H. Anticipation of negative pictures enhances the P2 and P3 in their later recognition. *Front. Hum. Neurosci.* **9**, 646 (2015).
54. Lin, H. *et al.* Expectation enhances event-related responses to affective stimuli. *Neurosci. Lett.* **522**, 123–127 (2012).
55. Nitschke, J. B., Sarinopoulos, I., Mackiewicz, K. L., Schaefer, H. S. & Davidson, R. J. Functional neuroanatomy of aversion and its anticipation. *NeuroImage* **29**, 106–116 (2006).
56. Ueda, K. *et al.* Brain activity during expectancy of emotional stimuli: an fMRI study. *NeuroReport* **14**, 51–55 (2003).
57. Onoda, K. *et al.* Anterior cingulate cortex modulates preparatory activation during certain anticipation of negative picture. *Neuropsychologia* **46**, 102–110 (2008).
58. van Pelt, S. *et al.* Beta- and gamma-band activity reflect predictive coding in the processing of causal events. *Soc. Cogn. Affect Neurosci.* **11**, 973–980 (2016).
59. Pfurtscheller, G., Stancák, A. & Neuper, Ch. Event-related synchronization (ERS) in the alpha band—an electrophysiological correlate of cortical idling: A review. *Int. J. Psychophysiol.* **24**, 39–46 (1996).
60. Foxe, J. J. & Snyder, A. C. The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Front. Psychol.* **2**, 2 (2011).
61. Jensen, O. & Mazaheri, A. Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Front. Hum. Neurosci.* **4**, 186 (2010).
62. Klimesch, W. α-band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci.* **16**, 606–617 (2012).
63. Klimesch, W., Sauseng, P. & Hanslmayr, S. EEG alpha oscillations: The inhibition-timing hypothesis. *Brain Res. Rev.* **53**, 63–88 (2007).
64. Pfurtscheller, G. Induced oscillations in the alpha band: Functional meaning. *Epilepsia* **44**, 2–8 (2003).
65. Uusberg, A., Uusberg, H., Kreegipuu, K. & Allik, J. EEG alpha and cortical inhibition in affective attention. *Int. J. Psychophysiol.* **89**, 2 (2013).
66. Chao, Z. C., Takaura, K., Wang, L., Fujii, N. & Dehaene, S. Large-scale cortical networks for hierarchical prediction and prediction error in the primate brain. *Neuron* **100**, 1252-1266.e3 (2018).
67. Engel, A. K. & Fries, P. Beta-band oscillations–signalling the status quo?. *Curr. Opin. Neurobiol.* **20**, 156–165 (2010).
68. Spitzer, B. & Haegens, S. Beyond the status quo: A role for beta oscillations in endogenous content (re)activation. *eNeuro* **4**, 2 (2017).
69. Spitzer, B., Gloel, M., Schmidt, T. T. & Blankenburg, F. Working memory coding of analog stimulus properties in the human prefrontal cortex. *Cereb. Cortex* **24**, 2229–2236 (2014).
70. Wimmer, K., Ramon, M., Pasternak, T. & Compte, A. Transitions between multiband oscillatory patterns characterize memory-guided perceptual decisions in prefrontal circuits. *J. Neurosci* **36**, 489–505 (2016).
71. Friston, K. Predictive coding, precision and synchrony. *Cogn. Neurosci.* **3**, 238–239 (2012).
72. Herrmann, C. S., Munk, M. H. J. & Engel, A. K. Cognitive functions of gamma-band activity: Memory match and utilization. *Trends Cogn. Sci.* **8**, 347–355 (2004).
73. Pearce, J. M. & Hall, G. A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol. Rev.* **87**, 532–552 (1980).


## Acknowledgements

We would like to thank Matthias Kerkemeyer for his help during data collection.

## Author contributions

A.S.: conceptualization, data curation, software, formal analysis, investigation, visualization, methodology, writing—original draft, project administration, writing—review and editing. M.R.: conceptualization, resources, software, methodology, writing—review and editing. S.F.: conceptualization, software, supervision. C.B.: conceptualization, resources, formal analysis, supervision, funding acquisition, validation, visualization, methodology, project administration, writing—review and editing.

## Funding

Open Access funding enabled and organized by Projekt DEAL. CB is supported by DFG SFB 289 project A02 and ERC-AdG-883892-PainPersist. MR is supported by DFG SFB 289 project A03 and DFG SFB TR 169 project B3. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project-ID 422744262–TRR 289. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. The study was approved by the Ethics board of the Hamburg Medical Association (PV4745). All methods were carried out in accordance with relevant guidelines and regulations. Informed consent was obtained from all participants.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-02939-z.

**Correspondence** and requests for materials should be addressed to A.S.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## 15.3 Study 3: Agency Affects Pain Inference through Intensity Shift as Opposed to Precision Modulation

Strube, A., Horing, B., Rose, M., & Büchel, C. (2022). Agency Affects Pain Inference through Intensity Shift as Opposed to Precision Modulation [Manuscript submitted for publication]. Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf

# Agency Affects Pain Inference through Intensity Shift as Opposed to Precision Modulation

Andreas Strube[1*], Björn Horing[1], Michael Rose[1] & Christian Büchel[1]

[1]Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany

Corresponding Author: Andreas Strube
Department of Systems Neuroscience
University Medical Center Hamburg-Eppendorf
20246 Hamburg, Germany
Phone: +49 40 7410 - 59899
email: a.strube@uke.de

Keywords: Agency, Expectation, Forward Model, Active Inference, Pain

## Abstract

Agency and expectations play a crucial role in pain perception and treatment. In the Bayesian pain model, somatosensation (likelihood) and expectations (prior) are weighted by their precision and integrated to form a pain percept (posterior). Combining pain treatment with stimulus-related expectations allows to mechanistically assess whether agency enters this model as a change in intensity or precision. In two experiments, heat pain was sham-treated either externally or by the subject, while a predictive cue was utilized to create high or low treatment expectations. Both experiments revealed additive effects and greater pain relief under self-treatment and high treatment expectations. Formal model comparisons favor models which allow intensity shifts rather than differences in precision. Electroencephalography revealed a theta-to-alpha effect associated with an interaction of expectations and agency, which was also correlated with trial-by-trial pain ratings. This effect was temporally associated with expectations, suggesting a shift regarding expectations (prior) rather than somatosensation (likelihood).


## Introduction

Tracey and Mantyh (2007) defined pain as "a conscious experience, an interpretation of the nociceptive input influenced by memories, emotional, pathological, genetic, and cognitive factors"[1]. In this context it has been shown that somatosensory processes such as pain perception are modulated by agency on a neurophysiological and behavioral level[2–16]. This beneficial effect is utilized in Patient-Controlled Analgesia (PCA) commonly used in post-operative care – patients receiving PCA experience less pain as compared to patients receiving traditional (i.e. externally applied) analgesia[17,18].

The somatosensory influence of agency on our perception is very intuitive - after all, we can experience that it is much harder to tickle oneself than to be tickled by another person, as has also been shown in a number of empirical studies[19–22]. Here, we utilize Bayesian modelling to assess the mechanisms how agency influence pain perception in the context of different pain treatment expectations.

It has been proposed that pain perception can be seen as a "Bayesian problem" requiring the integration of expectations with stimulus intensity[23–26]. The idea is that expectations are integrated with incoming nociceptive stimulus information, and both are weighted by their respective precision to form a pain percept. This has been shown by manipulation of the level of precision of prior treatment expectations, where expectation-based effects were more pronounced with more precise treatment expectations[27].

Similar to an application in visual perception[28], less precise sensory information would lead to a relatively higher influence of prior expectation, while more precise sensory information would lead to less influence of prior expectation on perception. Importantly, it is possible to design a pain treatment experiment in which sensory evidence (i.e. enhanced or reduced treatment efficacy) is either self-generated or externally generated.

Agency can now act at various points in the Bayesian pain model (see Fig. 1a). First, the influence of agency could occur via a shift in the *mean value* of the likelihood (we will term this likelihood shift model, see Fig. 1b) or a shift of the *mean value* of the prior. The former would entail that sensory neuronal processing is altered in intensity, whereas in the latter case, agency would change intensity expectations which influence the pain experience. Secondly, in the same manner, agency could change the *precision* of the likelihood (we will term this likelihood precision modulation model, see Fig. 1c) or the *precision* of the prior.

The influence of agency (i.e. self-generation of stimuli) on somatosensation is typically termed sensory attenuation - for which there are explanations comparable to the hypotheses derived from the Bayesian pain model. Charles Darwin (1872) already theorized about the influence of action on sensory precision: "from the fact that a child can hardly tickle itself, or in a much less degree than when tickled by another person, it seems that the precise point to be touched must not be known"[29].

To account for the effects of agency different models have been postulated. Firstly, there is the forward model which explains that smaller prediction errors during self-generated movement lead to a less intense sensation of action outcomes[30]. Translated to our pain treatment paradigm, this would represent a shift of the likelihood by self-treatment in the direction of experiencing less pain, which motivated our likelihood shift model (see Fig. 1b). Secondly, in the context of the active inference framework, sensory attenuation is discussed as necessary to enable action by lowering the precision of sensory evidence to the consequences of one's own actions[31]. This would translate to a reduction of precision of the likelihood in the Bayesian pain model, which motivated our likelihood precision modulation model (see Fig. 1c).

Derived from the Bayesian pain model of expectation-based hypoalgesia (see Fig. 1), we hypothesized that if the likelihood was shifted by agency, self-treatment would result in overall lower pain ratings (based on the forward model), regardless of prior expectations, i.e. agency and expectation effects would be additive. In contrast, less precision (based on active inference) of self-generated sensory consequences would enhance the impact of expectation effects (i.e. placebo/nocebo effects), leading to pain ratings being influenced more strongly by expectations in self-treatment conditions, which would manifest as an interaction of agency and expectations. See Fig. 1d for statistical hypotheses based on the

likelihood shift; see Fig. 1e for statistical hypotheses based on attenuated likelihood precision.

Consequently, we have translated this to formal models of Bayesian integration (Fig. 1a) in pain perception, incorporating a likelihood shift (Fig. 1b) and likelihood precision modulation (Fig. 1c), respectively. With these models, we performed a formal Bayesian model selection (see Fig. 1 for an overview of the main candidate models and model predictions). Please note that our models also consider the contrary mechanisms, i.e. the likelihood could be shifted in a way that self-treatment is associated with more pain. Also, the likelihood could become more precise leading to less influence of prior expectations. Bayesian modelling does not allow to disentangle whether the influence of agency is represented in a shift of the likelihood or the prior – a shift in the prior could equally account for the results as a shift in the likelihood. Differences in precision, on the other hand, would lead to different parameters for the posterior (i.e. in order to increase the effects of expectations, higher precision of the prior instead of a lower precision of the likelihood would lead to a higher precision of the posterior). Therefore, we also included a prior precision modulation model and models with multiple parameters (shift + precision modulation) for comparison.

To further measure the neural influence of agency, EEG can help to answer the question of whether agency acts on the likelihood or the prior in a Bayesian pain model. Modulations of EEG power in pain are typically more associated with signaling sensory information than with signaling expectations[32,33], and typically, expectation effects are related to processes occurring temporally before pain stimulation (e.g. after an expectation-inducing cue)[32,34]. An influence of agency on stimulus intensity-related EEG time-frequency patterns would therefore suggest a modulation of the likelihood, whereas

an influence of EEG time-frequency patterns related to expectation would suggest a modulation of the prior.

The first experiment ($N$ = 25) used continuous pain ratings to establish a precise readout of pain perception during painful heat stimulation and after treatment, while the second experiment ($N$ = 54) additionally employed electroencephalography (EEG) to further evaluate neurophysiological correlates of the modulation via expectations and agency. In both experiments, we applied heat pain to capsaicin-sensitized skin on the left radial forearm, after individual calibration to create comparable pain levels for each participant. To avoid a contamination of EEG data by movement artifacts through button presses during continuous pain ratings, we altered the paradigm for experiment 2 to include single outcome ratings instead of the continuous pain rating (see Fig. 2 for an overview of the trial design).

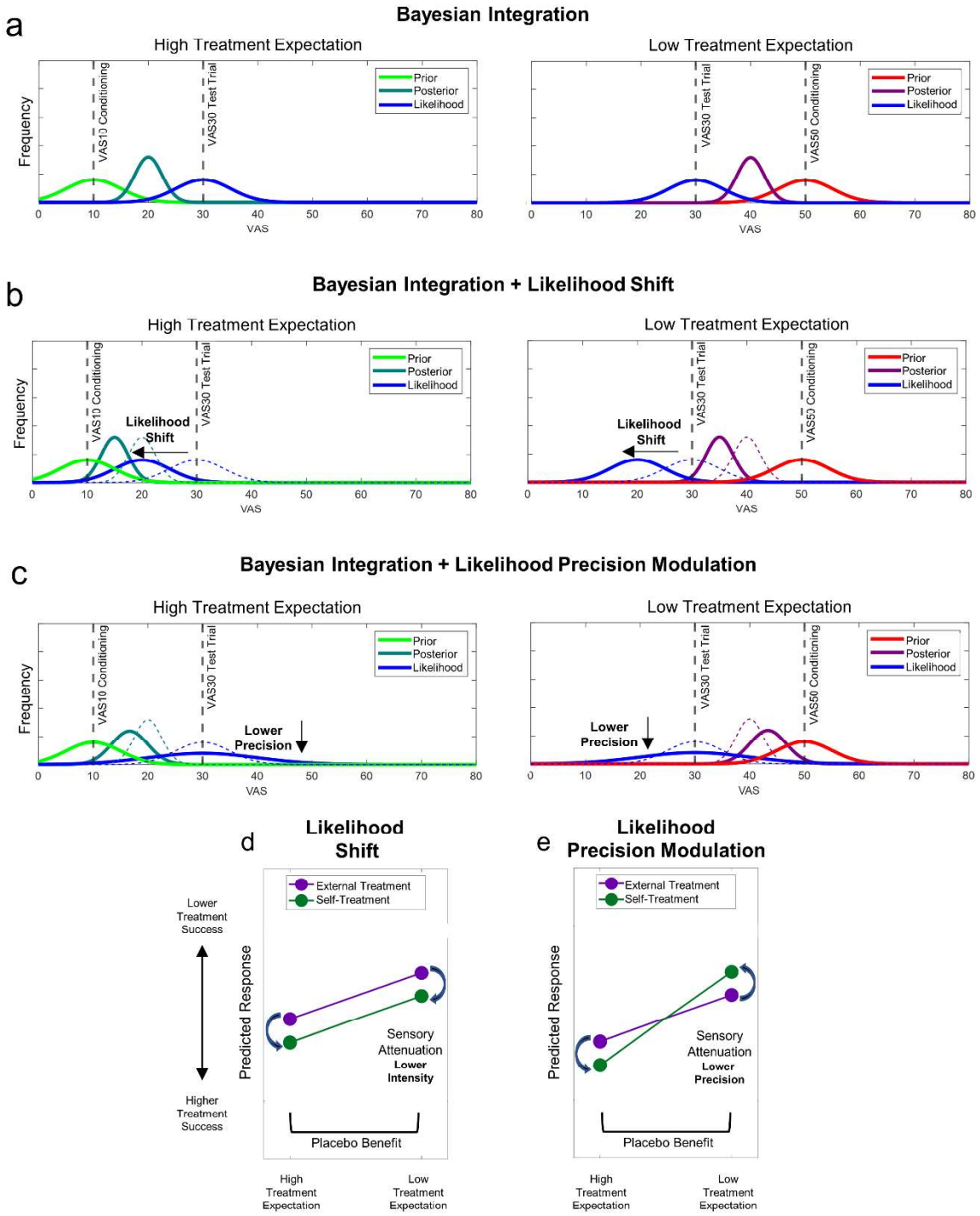



**Figure 1.** Bayesian models of pain perception. Bayesian model comparison was used to evaluate two main Bayesian pain placebo/nocebo models. (a) The core of both models is the Bayes-optimal integration of prior experiences (here centered at VAS = 10 for placebo and at VAS = 50 for nocebo) with incoming nociceptive information (i.e. likelihood) to form a pain percept (i.e. posterior). Prior, likelihood and posterior were approximated by Gaussian distributions allowing for an analytic solution of Bayesian integration. The likelihood shift model (b) has a free parameter that allows to shift the likelihood mean for self-treatment

trials (see STAR Methods, Eq.2). For example, the likelihood mean for self-treatment can be shifted to lower values (blue, solid line) as compared to the mean for external treatment (blue, dashed line). In high treatment expectation (placebo), this will lead to a shift of the posterior (dark green, solid line) to lower VAS values as compared to external treatment (dark green, dashed line) because of the integration of the shifted lower likelihood with the prior. Similarly, in low treatment expectation (nocebo), this will lead to a shift of the posterior (purple, solid line) to lower VAS values as compared to external treatment (purple, dashed line). This is in contrast to the likelihood precision modulation model (c) which has a free parameter that can change likelihood precision (see STAR Methods, Eq.3). If self-treatment is linked to a lower likelihood precision (blue, solid line) as compared to external treatment (blue, dashed line) this model should explain the data better than the likelihood shift model. As an example, this can lead to a posterior (dark green, solid line) which is more strongly drawn to the prior (VAS10 conditioning) in self-treatment than external treatment (dark green, dashed line), due to the lower "impact" of the likelihood. In low treatment expectation, the posterior (purple, solid line) would be drawn more strongly to the prior (VAS50 conditioning) in self-treatment than external treatment (purple, dashed line). Note that for actual modeling we utilized individual prior and likelihood parameters, whereas here, parameters are based on calibration target values for illustration purposes. The likelihood shift to lower values (d; derived from the forward model) predicts a decrease of perceived stimulus intensity in self-treatment (green line) as compared to external treatment (purple line), meaning a higher treatment success in self-treatment trials as compared to external treatment trials. Likelihood precision modulation leading to lower precision of the likelihood (e; derived from active inference) would entail that self-treatment is associated with a decrease in precision, and thus a larger influence of expectations in self-treatment (green line) as compared to external treatment (purple line).

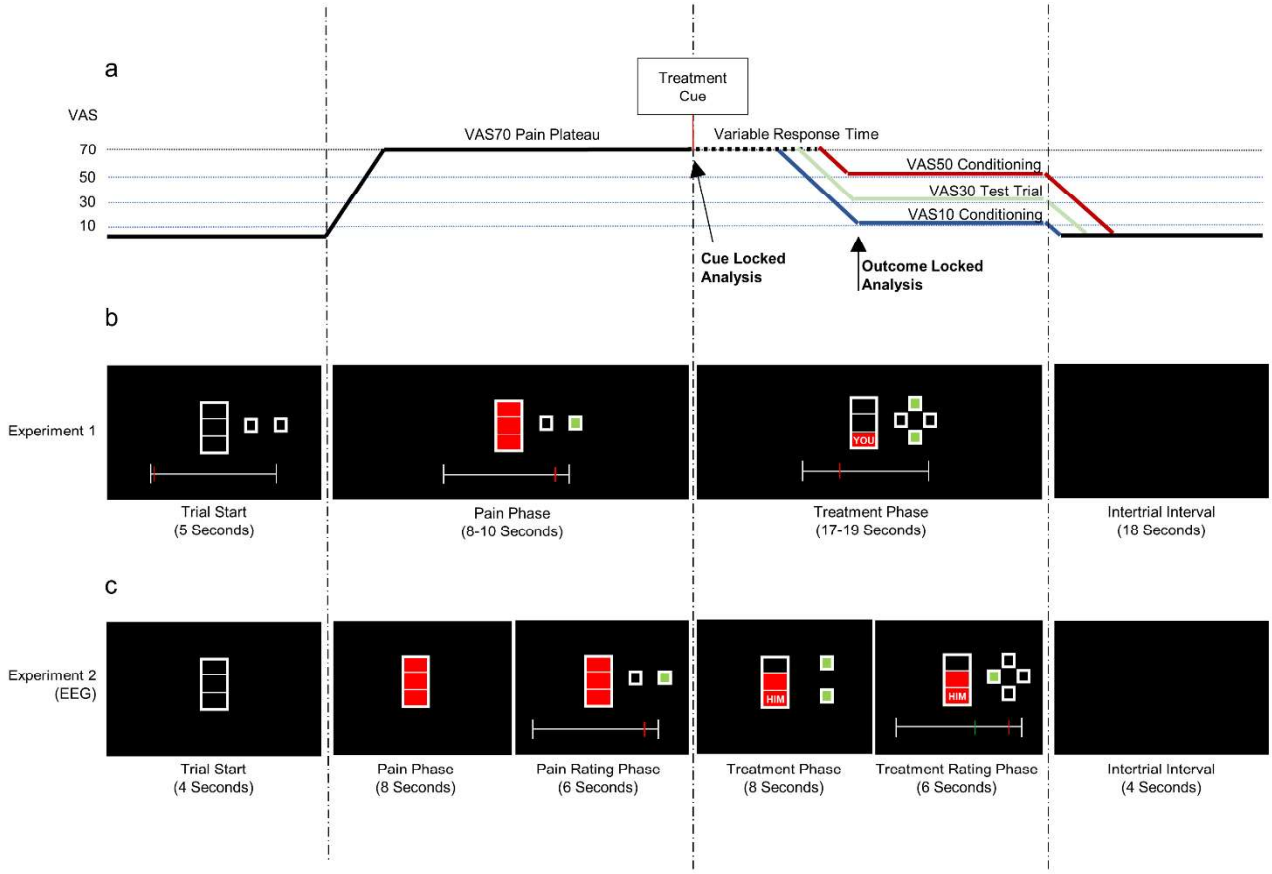


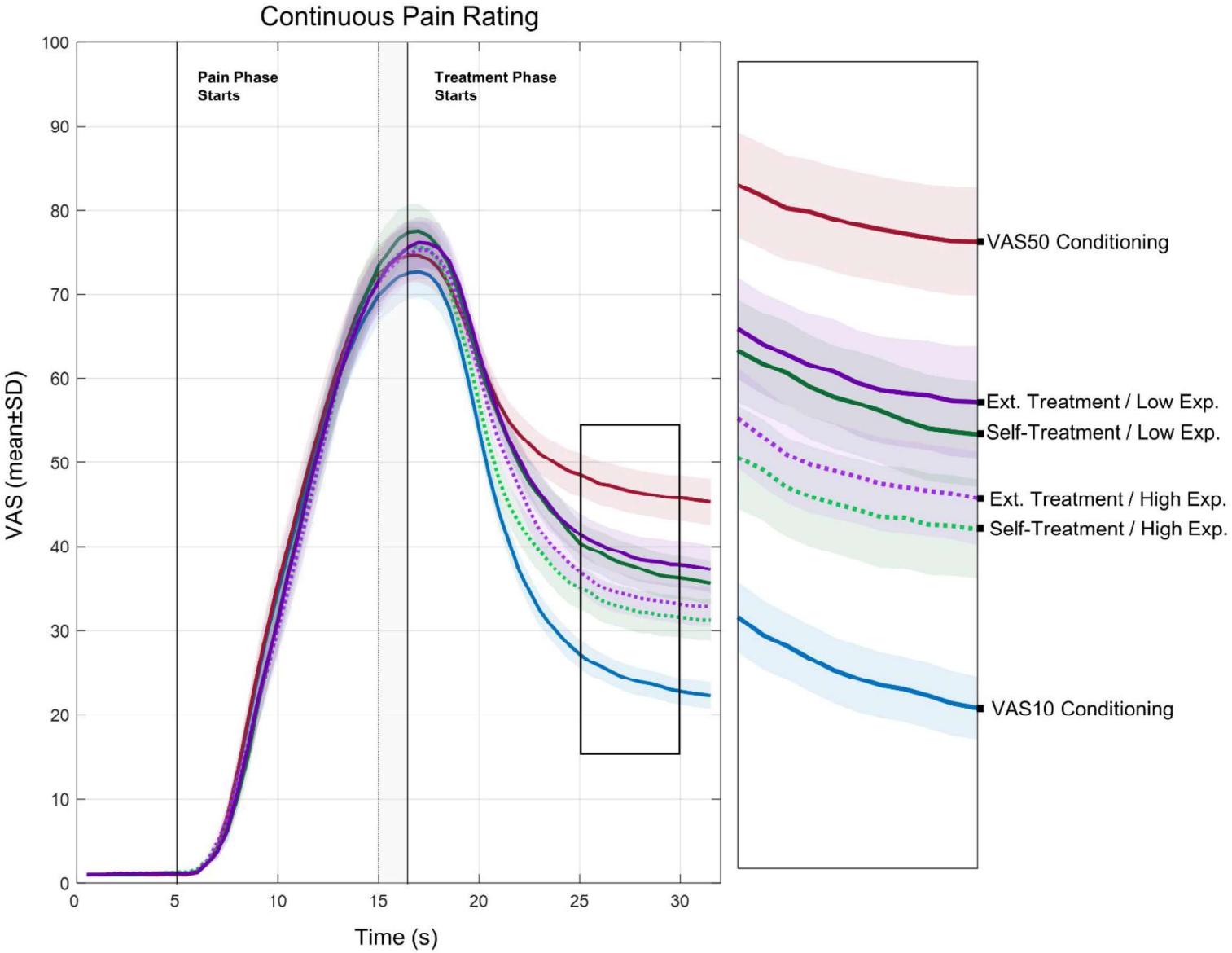

**Figure 2.** (a) Schematic representation of the paradigm, (b) trial design for experiment 1, and (c) trial design for experiment 2. Colored lines represent VAS50 conditioning (red), test trials (VAS30; green) and VAS10 conditioning (blue). The black line represents alterations in temperature common to all trial types, blue, green and red lines represent changes based on VAS10 conditioning, VAS30 test trials and VAS50 conditioning trials, respectively. At trial start, the thermal-heat stimulator (thermode), attached to the left radial forearm of the participant, is at the baseline temperature (set to 30°C for experiment 1 and to 28°C for experiment 2). A red bar indicates the start of the pain phase concurrent with an increase of thermode temperature to the individually calibrated pain level of VAS70. The start of the treatment phase is indicated by a cue showing whether self- or external treatment and whether highly or weakly effective treatment follows. This then leads to actual low (VAS10) or high (VAS50) temperatures during conditioning trials. In test trials, the final temperature is always at VAS30 regardless of the cued treatment efficacy. Arrows indicate time points for EEG data locks, i.e. the time axis of EEG time-frequency data was set to 0 according to the onset of the cue and to the treatment outcome (i.e. target treatment VAS level was reached by the thermode), respectively. In experiment 1 (b), a rating scale controlled with two buttons was presented during the whole trial. At trial start (5s), an empty bar was presented alongside the rating scale (set to 0 at the beginning) and a display of rating buttons (lighting up in green when pressed). During the following

pain phase (8-10s), the empty bar was filled red as an indication for pain. After the pain phase, the treatment cue was presented. The treatment cue showed a reduction of the red bar, where a reduction by 2/3 of the total height was associated with highly effective treatment and a reduction by 1/3 of the total height was associated with a weakly effective treatment. Additionally, a signal word indicated self- or external treatment. After a lag of 2s, the treatment buttons appeared on the display, lighting up in green when pressed by the subject or externally. After the treatment button was pressed, the temperature was decreased to the respective pain level. An ITI (intertrial interval) of 18s followed. In experiment 2 (c), pain ratings scales and buttons were only presented during designated rating phases. Treatment could be started immediately after the onset of the treatment cue. Here, the timing of each trial was: trial start (4s), pain phase (8s), pain rating phase (6s), treatment phase (8s) and treatment rating phase (6s) with an ITI of 4s.

# Results

## Experiment 1: Behavioral results

The first experiment used continuous pain ratings (Fig. 3). As we did not assess expectations explicitly we refer to expectations as the effects elicited by the predictive cue, i.e. high and low treatment expectation conditions are related to the predictive cues signaling high or low treatment success and not to actual expectation ratings.




184 **Figure 3.** Continuous VAS ratings per condition. Each line represents a different condition, i.e. VAS10 and
185 VAS50 conditioning, and four test conditions following VAS30 (self- versus external treatment, low versus
186 high treatment expectation). Pain phase (VAS70) starts after a cue presentation of 5s for a jittered duration
187 of 8-10s. Afterwards the treatment phase started, beginning with the presentation of the treatment cue for
188 2s. Then, treatment was started either by the participant or externally. Lines on the right represent an
189 enlargement of the highlighted section (25-30s).

190 The treatment outcome differed objectively as we reduced the pain stimulus to three
191 different intensities, i.e. VAS10 and VAS50 for high and low treatment efficacy
192 respectively during conditioning trials, and VAS30 for test trials (albeit presented with

the respective high or low conditioned predictive cues). To evaluate if participants experienced these three intensities to be different, we conducted a repeated measures ANOVA on the final continuous rating data points (post-treatment VAS rating) from all three stimulus intensities, including conditioning and test trials (averaging across predictive cues) which revealed a significant difference ($F(2,48) = 43.78$, $p < 0.001$, $\eta_p^2 = 0.646$) (Fig. 4a).

Post-hoc analyses using Bonferroni correction for multiple comparisons indicated that all three stimulus intensity levels differed significantly from each other, revealing higher post-treatment VAS ratings for VAS50 conditioning trials ($M = 50.35$, $SD = 15.78$) versus VAS30 test trials ($M = 38.98$, $SD = 12.86$) and for VAS30 test trials versus VAS10 conditioning trials ($M = 24.47$, $SD = 11.95$; all $p < 0.001$).

As a next step, we evaluated the effects of our manipulations for the test trials, where the intensity of the painful stimulus was always reduced to an individually calibrated level of VAS30. Here, post-treatment VAS ratings could either be influenced by agency (self- versus external treatment), expectations (low versus high treatment expectations), or their interaction. Considering the likelihood shift model (see Fig. 1d), we would expect higher treatment success in self-treatment test trials as compared to external treatment test trials regardless of prior treatment expectations (as derived from the forward model). In other words, we would expect a main effect of agency with or without a main effect of expectation, but no interaction between both factors. Conversely, for the likelihood precision modulation model (see Fig. 1e), by lowering the precision of sensory evidence (as derived from predictions related to active inference), prior treatment expectations would gain a higher relative weight compared to the sensory information in self-treatment trials versus external treatment trials, which would manifest as an interaction

(i.e. expectation effects should be larger in self-treatment). Note that self-treatment could also be associated with worse outcomes and therefore a shift of the likelihood to higher VAS values. Also, self-treatment could lead to less influence of expectations and a more precise likelihood in the likelihood precision modulation model.

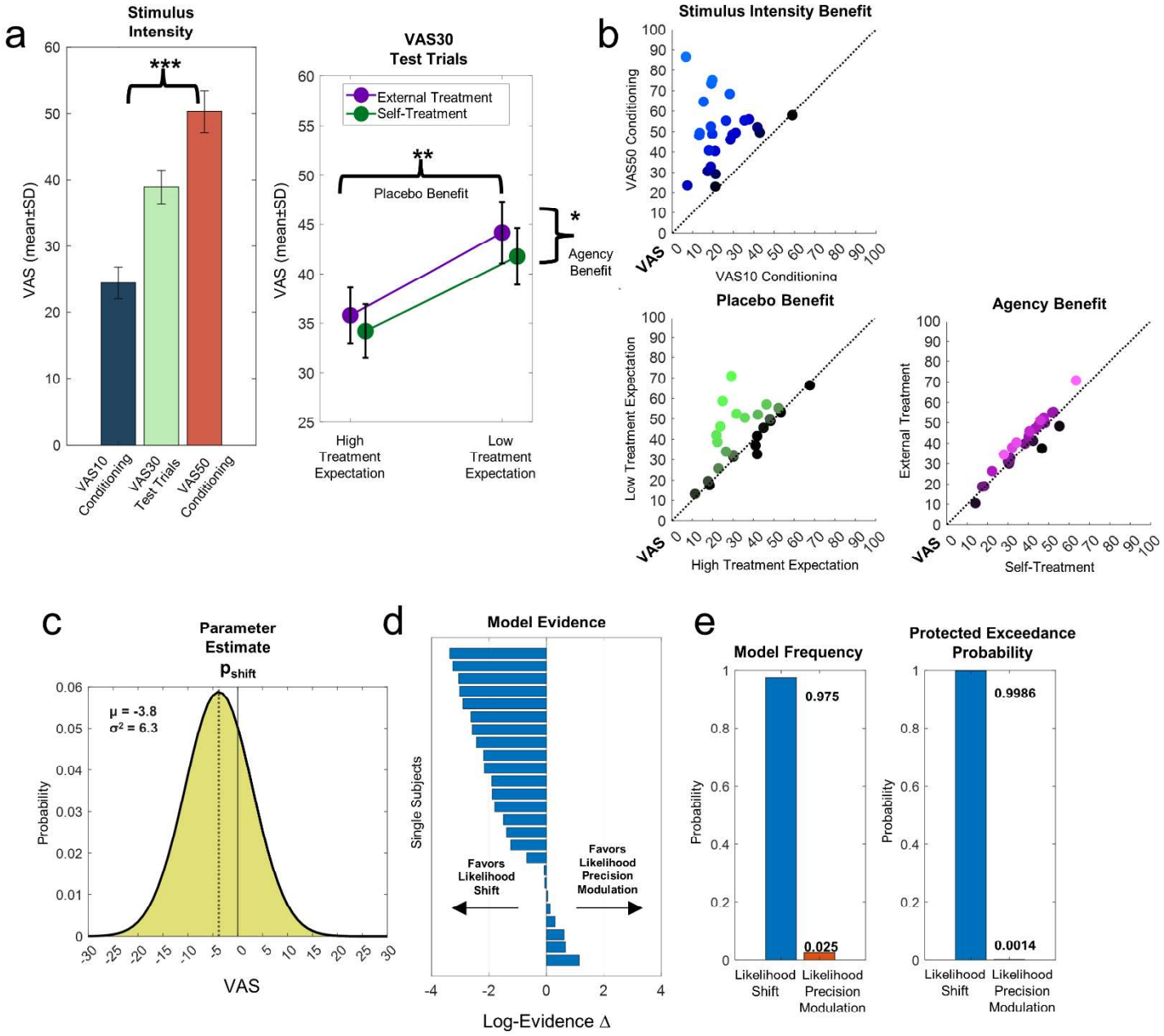
Here, again we conducted a repeated measures ANOVA to test for main effects of agency, expectation, and their interaction in the test conditions (VAS30). We found a significant sensory attenuation effect, that is, a main effect of agency ($F(1,24) = 6.2$, $p = 0.02$, $\eta_p^2 = 0.205$), meaning that post-treatment VAS ratings were lower for self-treatment trials ($M = 37.99$, $SD = 12.57$) versus external treatment trials ($M = 39.98$, $SD = 13.45$). Furthermore, we found a significant expectation effect ($F(1,24) = 10.738$, $p = 0.003$, $\eta_p^2 = 0.309$), i.e. high treatment expectations were associated with lower post-treatment VAS ratings ($M = 35.00$, $SD = 13.66$; conditioned with VAS10) than those following low treatment expectations ($M = 42.97$, $SD = 14.77$; conditioned with VAS50). Importantly, we did not observe a significant interaction of expectation and agency ($F(1,24) = 0.679$, $p = 0.42$, $\eta_p^2 = 0.028$). A linear regression of reaction times and agency benefits (self-treatment minus external treatment post-treatment VAS rating) did not reveal a significant association of these factors ($R^2 = 0.00503$, $F(1,24) = 0.116$, $p = 0.736$).

For model-based analyses of our behavioral data, we created two Bayesian models of pain perception in placebo/nocebo pain treatment (see Fig. 1) which were inverted and compared using variational Bayesian methods (VBA, see STAR Methods for details). We used the Bayesian integration model of pain perception as a basis model (Fig. 1a). In self-treatment test trials under the likelihood shift model, we included the parameter $p_{shift}$ which allowed for a shift of the likelihood and thus for a posterior distribution which was shifted into the same direction in low and high treatment expectation conditions (Fig. 1b;

here a shift of the likelihood mean to lower values leads to better treatment outcomes). This was contrasted to the likelihood precision modulation model, where the posterior distribution should be differentially affected by the conditioned pain experience, and thus we included the parameter $p_{precision}$ which allowed for increases and decreases in likelihood precision (see Fig. 1c, here a relaxation of likelihood precision leads to an increased relative weight of expectations).

We used a random effects (RFX) Bayesian model selection approach[35,36] to estimate the overall posterior model probability across subjects. The RFX model exceedance probability was at $\varphi = 0.9986$ for the likelihood shift model compared to $\varphi = 0.0014$ for the likelihood precision modulation model. Hence, we see clear evidence for the likelihood shift model with a free parameter enabling likelihood shift over the likelihood precision modulation model with a free parameter allowing a modulation of likelihood variance (see Fig. 4e). The likelihood shift model also outperforms all other control models in a full comparison, i.e. a null model without free parameters (i.e. setting the likelihood shift parameter to 0 and the likelihood precision modulation parameter to 1) and a full model which included both shift and precision modulation parameters ($\varphi > 0.999$) (see STAR Methods for details). See Supplementary Fig. 1 for a comparison of all candidate models.

In summary, results using continuous pain ratings clearly demonstrate both sensory attenuation effects (i.e. self-treatment lead to better outcomes) and expectation effects (i.e. high treatment expectations lead to better treatment outcomes), but no interaction between expectation and agency. Model selection provides strong evidence in favor of the likelihood shift model over all other candidate models (see Fig. 4 for a summary of the results).




264

265 **Figure 4.** Results from VAS (Visual Analogue Scale) rating analyses of experiment 1 (*N* = 25). (a) Post-
266 treatment VAS ratings for each stimulus intensity condition (VAS10, 30 and 50) and lines representing the
267 contrasts of low versus high treatment expectation and self- versus external treatment during test trials.
268 Bars and lines represent post-treatment VAS ratings averaged per condition. (b) Scatter plots represent
269 single subject values for treatment outcomes for conditioning, expectation, and agency. Scatter plots
270 represent contrasts of conditions, i.e. each dot represents averaged ratings of a single subject for VAS10

versus VAS50 conditioning (blue), high treatment expectation versus low treatment expectation (green) and self-treatment versus external treatment (purple). Brighter colors indicate larger benefits of stimulus intensity (VAS10 versus VAS50 conditioning), placebo benefits (high treatment expectations versus low treatment expectations) and agency benefits (self-treatment versus external treatment). Data points above the diagonal represent single subjects with stimulus intensity, placebo and agency benefits, respectively. (c) Probability density function of group parameter estimates for the likelihood shift parameter $p_{shift}$ of the winning likelihood shift model. (d) Single subject differences of log evidence for the likelihood shift model versus likelihood precision modulation model (negative values favor the likelihood shift model) and (e) model frequencies and protected exceedance probabilities.

## Experiment 2: Behavioral results

As in the first experiment, a repeated measures ANOVA with all three stimulus intensities (including conditioning and test trials) revealed a significant difference ($F(2,106) = 118.32$, $p < 0.001$, $\eta_p^2 = 0.691$) between all three intensities (VAS10, 30 and 50) in post-treatment VAS ratings. Post-hoc analyses using Bonferroni correction for multiple comparisons indicated that all three stimulus intensity levels differed significantly from each other, revealing higher post-treatment VAS ratings for VAS50 conditioning trials ($M = 49.66$, $SD = 15.46$) versus VAS30 test trials ($M = 38.24$, $SD = 17.18$) and for VAS30 test trials versus VAS10 conditioning trials ($M = 31.04$, $SD = 17.45$; all $p < 0.001$).

For the evaluation of main effects of agency and expectation and their interaction for the test trials, we again conducted a repeated measures ANOVA. We observed a significant sensory attenuation effect ($F(1,53) = 19.13$, $p < 0.001$, $\eta_p^2 = 0.265$), i.e. post-treatment VAS ratings were lower for self-treatment trials ($M = 37.15$, $SD = 17.26$) as compared to external treatment trials ($M = 39.33$, $SD = 17.30$). Also, we found a significant expectation effect ($F(1,53) = 35.57$, $p < 0.001$, $\eta_p^2 = 0.402$), i.e. high treatment expectations were

associated with lower post-treatment VAS ratings ($M$ = 34.91, $SD$ = 17.69; conditioned with VAS10) than low treatment expectations ($M$ = 41.56, $SD$ = 17.64; conditioned with VAS50). As in the first experiment, we did not observe a significant interaction of treatment expectation and agency ($F(1,53) = 0.02$, $p = 0.887$, $\eta_p^2 = 0.003$). Finally, we assessed if agency effects might be explained by reaction times. We tested if reaction times significantly explained agency benefits (self-treatment minus external treatment post-treatment VAS rating) in a linear regression model, which showed no significant effect ($R^2 = 0.0286$, $F(1,52) = 1.53$, $p = 0.221$).

Again, we used a random effects (RFX) Bayesian model selection approach to estimate the overall posterior model probability across subjects for the post-treatment VAS ratings in experiment 2. For experiment 2, the RFX exceedance probability of $\varphi = 0.9995$ for the likelihood shift model compared to $\varphi = 0.0005$ for the likelihood precision modulation model again strongly favored the likelihood shift model over the likelihood precision modulation model (Fig. 5e). The likelihood shift model also wins against all other candidate models ($\varphi > 0.999$) in a full comparison (see STAR Methods for details). See Supplementary Fig. 1 for a comparison of all candidate models.

Taken together, in the second experiment, we replicated the rating-related results of the first experiment. Again, these results demonstrate sensory attenuation effects (i.e. self-treatment leads to better treatment outcomes) and expectation effects (i.e. high treatment expectations lead to better treatment outcomes) but no interaction between expectation and agency. Model selection again provides strong evidence in favor of the likelihood shift model over the likelihood precision modulation model (see Fig. 5 for a summary of the results).

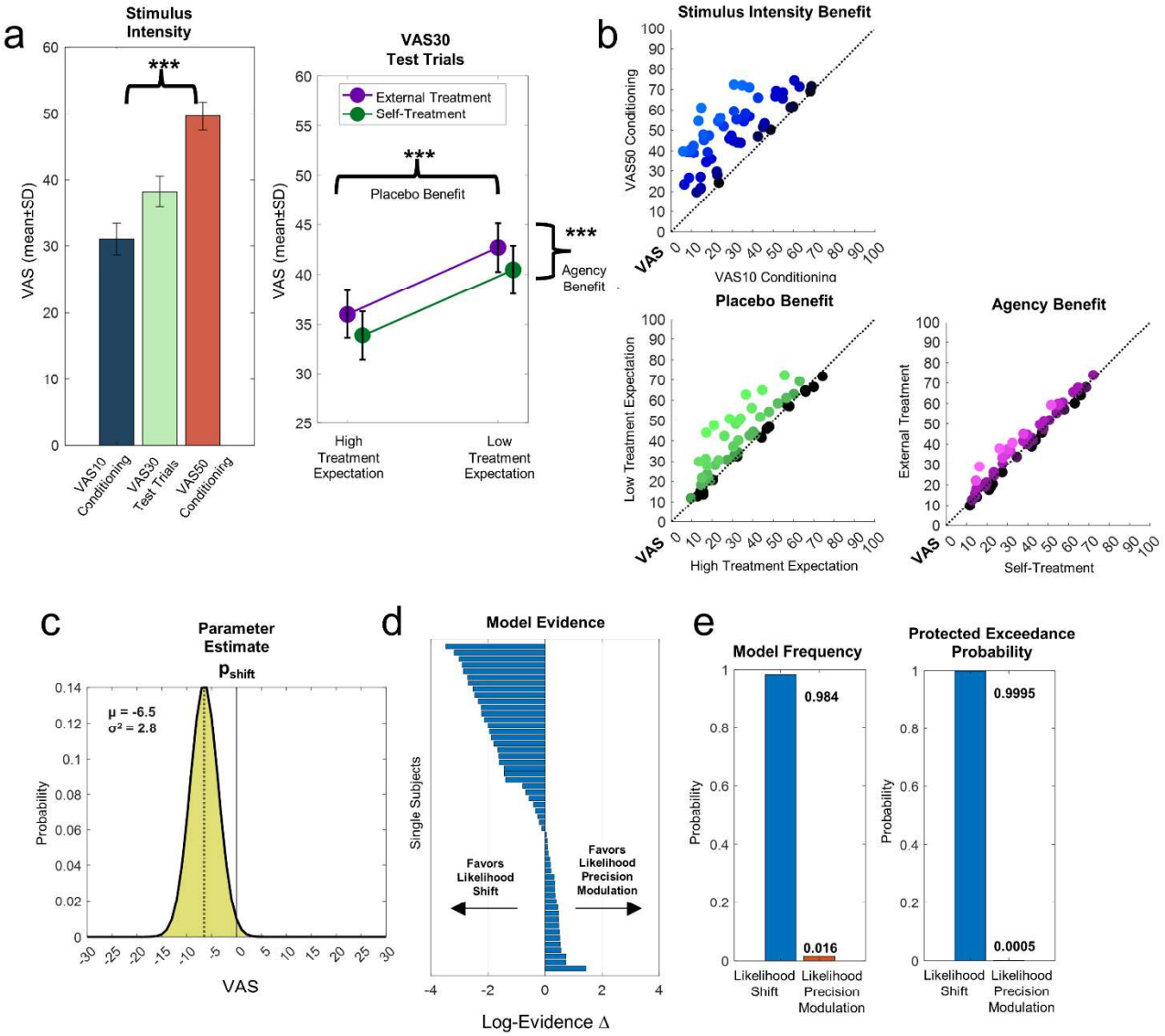

a
Stimulus Intensity
VAS (mean±SD)
VAS10 Conditioning
VAS30 Test Trials
VAS50 Conditioning
***
VAS30 Test Trials
External Treatment
Self-Treatment
Placebo Benefit
Agency Benefit
High Treatment Expectation
Low Treatment Expectation
b
Stimulus Intensity Benefit
VAS50 Conditioning
VAS10 Conditioning
Placebo Benefit
Low Treatment Expectation
High Treatment Expectation
Agency Benefit
External Treatment
Self-Treatment
c
Parameter Estimate
p_shift
μ = -6.5
σ² = 2.8
Probability
VAS
d
Model Evidence
Single Subjects
Favors Likelihood Shift
Favors Likelihood Precision Modulation
Log-Evidence Δ
e
Model Frequency
0.984
0.016
Likelihood Shift
Likelihood Precision Modulation
Protected Exceedance Probability
0.9995
0.0005
f
Bayesian Integration
Empirical VAS Ratings
Theoretical Posterior
Prior
Posterior
Likelihood
Frequency
VAS
Bayesian Integration + Likelihood Shift
Likelihood Shift
High Treatment Expectation
Self-Treatment
Low Treatment Expectation
Self-Treatment

**Figure 5.** Behavioral VAS (Visual Analogue Scale) post-treatment pain rating data of experiment 2 ($N$ = 54). (a) Post-treatment VAS ratings for each stimulus intensity condition (VAS10, 30 and 50) and lines representing the contrasts of low versus high treatment expectation and self- versus external treatment during test trials. Bars and lines represent post-treatment VAS ratings averaged per condition. (b) Scatter plots represent single subject values for treatment outcomes for conditioning, expectation, and agency. Scatter plots represent contrasts of conditions, i.e. each dot represents averaged ratings of a single subject for VAS10 versus VAS50 conditioning (blue), high treatment expectation versus low treatment expectation (green) and self-treatment versus external treatment (purple). Brighter colors indicate larger benefits of stimulus intensity (VAS10 versus VAS50 conditioning), placebo benefits (high treatment expectations versus low treatment expectations) and agency benefits (self-treatment versus external treatment). Data points above the diagonal represent single subjects with stimulus intensity, placebo and agency benefits, respectively. (c) Probability density function of group parameter estimates for the likelihood shift parameter $p_{shift}$ of the winning likelihood shift model. (d) Single subject differences of log evidence for the likelihood shift versus likelihood precision modulation model (negative values favor the likelihood shift model) and (e) model frequencies and protected exceedance probabilities. (f) Example data of a single subject clearly indicates an improved fit of behavioral post-treatment VAS ratings with the likelihood shift model (right) over basic Bayesian integration (left). Self-treatment conditions with high and low treatment expectations (left) with a theoretical posterior based on basic Bayesian integration and (right) with a theoretical posterior based on Bayesian integration with a likelihood shift (i.e. sensory attenuation motivated by the forward model) are shown. Lines represent empirical Gaussian high treatment expectation priors (green), low treatment expectation priors (red) and likelihood (blue) based on the respective fitted model. The solid dark green line represents the theoretical posterior based on Bayesian integration of the high treatment expectation prior and the likelihood. The dashed dark green line represents the parameters of a fitted Gaussian distribution to the empirical post-treatment VAS ratings of the respective conditions. Accordingly, the solid purple line represents the theoretical posterior based on Bayesian integration of the low treatment expectation prior and the likelihood. The dashed purple line represents the parameters of a fitted Gaussian distribution to the empirical post-treatment VAS ratings of the respective conditions.

## Experiment 2: EEG time-frequency data

For the statistical analysis of EEG data, we considered two separate time points for time-frequency data to evaluate cue-locked as well as treatment outcome-locked effects. For cue-locked analyses, we set $t = 0$ to the onset of the cue indicating the conditioned effectiveness of the treatment and the agency condition of the treatment phase. In outcome-locked analyses $t = 0$ was set to the point when the thermode reached the calibrated treatment VAS target, and thus, takes individual variations in treatment latency into account. All tests were corrected for multiple comparisons using Monte Carlo cluster tests. At each sample a t-test was conducted for each respective contrast (i.e. conditioning, expectation, agency and interaction) and all samples exceeding the threshold of $p < .05$ were clustered in connected sets on the basis of temporal (i.e. adjacent time points), spatial (i.e. neighboring electrodes), and spectral adjacency. This was repeated with shuffled condition labels per subject and the cluster p-value is calculated as the proportion of clusters exceeding the original cluster masses (i.e. sum of all *t*-values at all samples within a cluster) in random permutation (see STAR Methods for details). See Supplementary Fig. 2, 3 and 4 for *z*-scored cue-locked time-frequency data and Supplementary Fig. 5, 6 and 7 for *z*-scored outcome-locked time-frequency data for all conditions at Fz, Cz and Pz, respectively.

## Conditioning

In each trial, a predictive cue indicated an upcoming highly or weakly effective treatment on the ongoing VAS70 stimulus. This association was established during conditioning trials, where a predictive cue indicating high treatment efficacy was associated with a (physical) stimulus intensity decrease to a temperature individually representing VAS10. Similarly, a cue indicating low treatment efficacy was associated

with a decrease to a temperature representing VAS50. A cluster-corrected dependent samples *t*-test on cue-locked data (0 to 2s after the cue, 4-181Hz) revealed no differences between VAS10 and VAS50 conditioning trials immediately after cue presentation (all $p > .05$). However, differences were significant for outcome-locked data (in a window of -1 to 2s with $t = 0$ at target temperature, 4-181Hz), revealing two clusters of activity associated with different conditioning types (VAS10 versus VAS50; Fig. 6; also see Supplementary Fig. 8 for topographies of significant clusters of activity for different frequency bands and Supplementary Fig. 9 for bar graphs representing the averaged power at significant clusters of activity for all conditions). Negative times reflect activity during the unfolding of pain relief, whereas positive times indicate activity during the outcome phase, i.e. when temperatures were stable at the final outcome level. We observed a positive cluster ($p < .001$) including frequencies from 8 to 64Hz in a time frame from -0.4 to 1s, indicating an increase of EEG power for VAS10 conditioning versus VAS50 conditioning. Additionally, we observed one negative cluster ($p < .001$) associated with decreased EEG power of VAS10 versus VAS50 conditioning. This cluster included frequencies from 4 to 45Hz in a time frame from -0.95 to 0.95s.

Longer outcome latencies based on increased ramp times in the VAS10 as compared to the VAS50 condition might explain differences in EEG data due to habituation, i.e. the target temperature is reached at later time points in a decrease from VAS70 to VAS10 as compared to VAS70 to VAS50. For example, a decrease to a VAS50 temperature of 43°C from a VAS70 plateau of 47°C takes 500ms whereas a decrease to a VAS10 temperature of 41°C takes 750ms. Therefore, we tested if activity in clusters representing differences in intensity (VAS10 versus VAS50 conditioning trials) were associated with ramp time differences between VAS10 and VAS50 conditioning trials. Both clusters were not associated with differences in outcome latencies in a linear regression analysis, i.e. the

individual differences in outcome latency based on different target temperatures in VAS50 and VAS10 conditions were not predictive for EEG power in the positive and negative clusters (all $p > .05$; see Supplementary Figure 10 for details).

This demonstrates a modulation of theta (4-8Hz), alpha-to-beta (8-30Hz) and low gamma (30-50Hz) frequencies by treatment efficacy via differences in stimulus intensity, i.e. more effective treatment (VAS10) was associated with lower alpha-to-beta (8-30Hz) activity during the relief phase followed by lower theta (4-8Hz) and increased alpha-to-beta (8-30Hz) power at the treatment outcome phase, as compared to less effective treatment (VAS50).

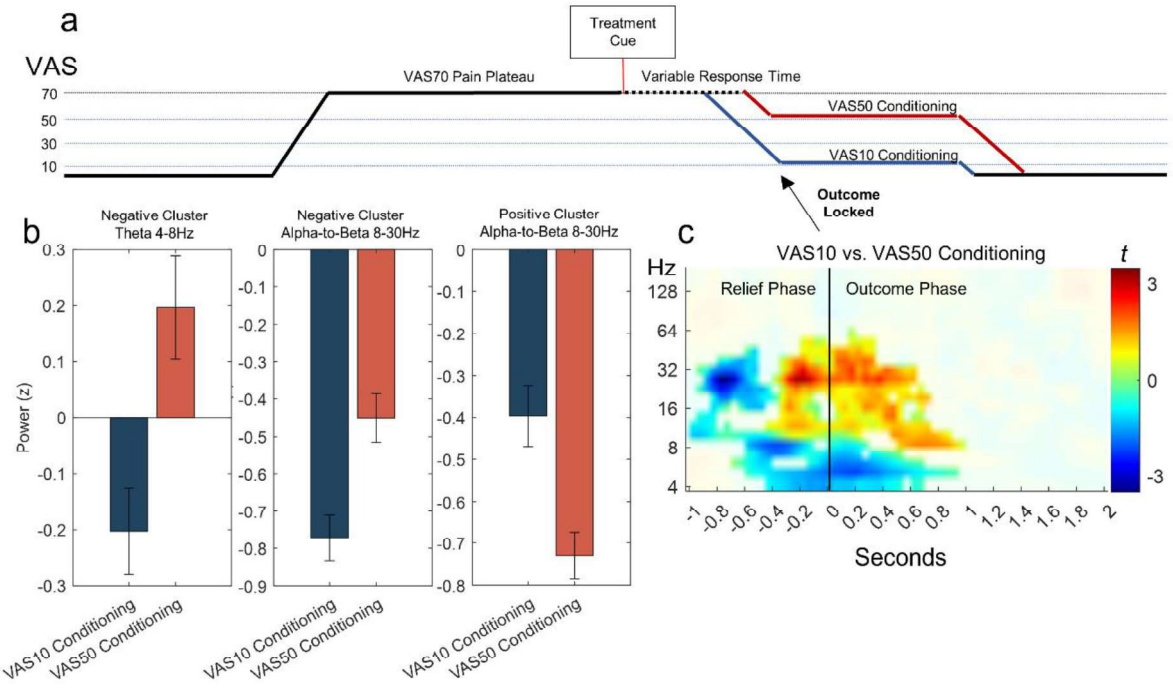



**Figure 6.** (a) Pain treatment paradigm, (b) differences in EEG power ($z$-scored) for VAS10 conditioning and VAS50 conditioning averaged over all significant samples in the theta (4-8Hz) and alpha-to-beta (8-30Hz) bands, (c) time-frequency plot of significant clusters of activity associated with the outcome-locked main effects of VAS10 versus VAS50 conditioning. Time-frequency plots are averaged $t$-values over all channels including significant data points of the respective clusters of activity. Non-significant data points are masked out. Colors represent $t$-values.

## Agency

Agency was experimentally manipulated as participants had to either initiate the treatment themselves, or the treatment was initiated (putatively) by the experimenter. To test for the effects of agency, we again considered both phases and conducted *t*-tests contrasting self-treatment and external treatment during test trials. Here, a cluster-corrected dependent samples t-test revealed no association between EEG time-frequency data and agency with cue-locked data (0 to 2s after cue onset, 4-181Hz; all $p > .05$). However, treatment outcome-locked data (-1 to 2s at target temperature, 4-181Hz) revealed two significant clusters of activity associated with agency (Fig. 7). Data showed a negative cluster ($p < 0.001$) ranging from -1 to 1s including frequencies from 4 to 54Hz, exhibiting a negative association between agency and EEG power during the relief (-1 to 0s) and outcome phase (0 to 2s). We also observed a significant ($p = .028$) positive cluster, ranging from -0.2 to 0.75s including frequencies from 11 to 54Hz, indicating increased alpha-to-beta (8-30Hz) and low gamma (30-50Hz) power during the treatment outcome phase for self-treatment versus external treatment.

Here, again, EEG activity might be explained by longer pain plateau durations which might lead to habituation. In this case, ramp durations were always identical (i.e. from VAS70 to VAS30), but in contrast, self-treatment is individually associated with slight differences in outcome latencies due to individual reaction times and thus potentially a longer pain plateau duration. Therefore, we assessed if variance of clusters associated with agency might also be explained by reaction times. We tested differences in self-treatment versus external treatment averaged EEG power over all samples included in each cluster in a linear regression model with the average reaction time of each participant as a regressor. This regressor explained a significant amount of the variance in the negative agency

cluster ($R^2$ = 0.228, $F$(1,52) = 15.4, $p$ < .001), but not in the positive cluster ($R^2$ = 0.0611, $F$(1,52) = 3.38, $p$ = 0.072). When partialling out the variance explained, the difference in EEG power in the negative clusters were not significant anymore (p > .05), whereas for the positive cluster this was not the case and the effect of agency remained significant ($t$(53) = 2.93; $p$ = 0.005). See Supplementary Figure 11 for further details on the association of reaction time data and agency clusters.

To further evaluate the influence of between-subject differences in sensory attenuation, we conducted a Pearson's correlation analysis based on the individually estimated likelihood shift parameter (i.e. agency benefit; see STAR Methods for details). Here, no significant associations between subject differences in likelihood shift were associated with EEG time-frequency data (all $p$ > .05) in cue-locked (0 to 2s after cue onset, 4-181Hz) and treatment outcome-locked (-1 to 2s at target temperature, 4-181Hz) data.

Finally, we assessed in a linear mixed effects model if the agency-related positive cluster was associated with individual trial-by-trial outcome pain ratings of the test trials (all VAS30; see STAR Methods). Activity at the positive agency cluster was not associated with VAS outcome ratings on a trial-by-trial basis for VAS30 test trials ($t$(2588) = -1.148, $p$ = .25). See Supplementary Table 2 for the full linear mixed effects model.

Taken together, the analysis of agency revealed two clusters associated with differences between self- and external treatment. Firstly, we observed a negative cluster during the relief and treatment outcome phase including frequencies from the theta (4-8Hz) to low gamma (30-50Hz) range, which was related to individual differences in reaction times. Secondly, when the target temperature was reached, data showed an increase of alpha-to-beta activity associated with self-treatment (as compared to external treatment). This cluster was not significantly associated with reaction times.

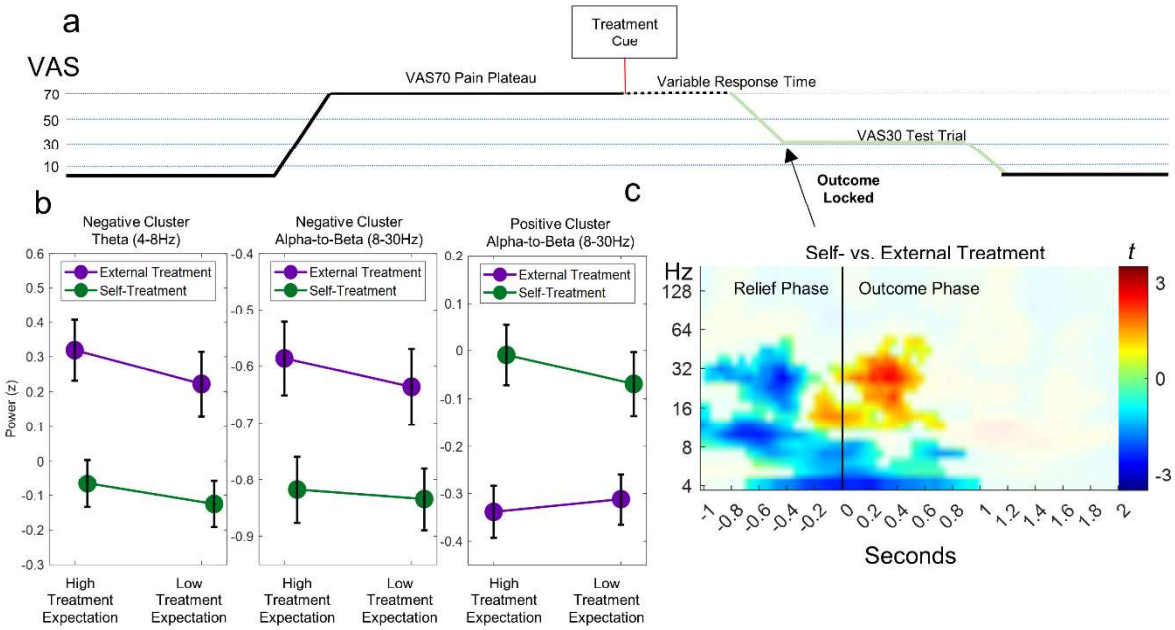



**Figure 7.** (a) Pain treatment paradigm, (b) a panel of line graphs representing differences in EEG power (*z*-scored) for self-treatment (green) versus external treatment (purple) averaged over all significant samples in the theta (4-8Hz) and alpha-to-beta (8-30Hz) ranges, and (c) a time-frequency plot representing significant clusters of activity associated with the outcome-locked main effects of self- versus external treatment. Time-frequency plots are averaged t-values over all channels including significant data points of the respective clusters of activity. Non-significant data points are masked out. Colors represent *t*-values.

## Expectation

As a next step, we evaluated the effects of treatment expectations (i.e. contrast of predictive cues). During test trials, stimulus intensity (i.e. the target temperature of the treatment) was set to VAS30. To evaluate the effects of different treatment expectations, we conducted a cluster-corrected dependent samples *t*-test on cue-locked data (0 to 2s after cue onset, 4-181Hz) which revealed no differences between high and low treatment expectations (all $p > .05$). Likewise, a cluster-corrected dependent samples *t*-test did not reveal significant clusters associated with treatment expectations at the treatment outcome (-1 to 2s at target temperature, 4-181Hz; all $p > .05$).

To further evaluate the effects of treatment expectations, we conducted a Pearson's correlation analysis on *z*-standardized behavioral expectation effects per subject (i.e. placebo benefit; see STAR Methods for details). Here, positive or negative clusters would indicate a correlative association of EEG power and the size of the expectation effects. However, there was no significant association of between-subject expectation effects and EEG time-frequency data (all $p > .05$) in cue-locked (0 to 2s after cue onset, 4-181Hz) and treatment outcome-locked (-1 to 2s at target temperature, 4-181Hz) data.

## Interaction of agency and expectation

Finally, we tested for an interaction of treatment expectation and agency. At cue-locked data (0 to 2s after cue onset, 4-181Hz), a cluster-corrected dependent samples *t*-test revealed a significant association of EEG power and the interaction term ($p = .034$; Fig. 8) ranging from 0 to 1.35s after cue onset and including frequencies from 4 to 13.5Hz. Here, we also conducted post-hoc t-tests, which revealed a crossed interaction (all $p < .05$, see Supplementary Data for detailed post-hoc *t*-test results confirming the crossed interaction). Treatment outcome-locked data did not reveal any cluster associated with an interaction of treatment expectation and agency (all $p > .05$). Activity at the interaction cluster was not associated with differences in averaged individual reaction times for self-treatment ($R^2 = 0.0505$, $F(1,52) = 2.76$, $p = 0.102$; see Supplementary Figure 12).

Again, we assessed in a linear mixed effects model if the interaction-related negative cluster was associated with individual trial-by-trial outcome pain ratings of the test trials (all VAS30; see STAR Methods). Average activity in the negative cluster was significantly associated ($\beta = 0.74$) with VAS outcome ratings on a trial-by-trial basis for VAS30 test trials ($t(2588) = 2.51$, $p = .0121$).

The analysis of the interaction term revealed differential integration of treatment expectation and agency information in theta (4-8Hz) and alpha (8-12Hz) frequencies starting shortly after the presentation of the cue (see Fig. 8 for a summary of the results). Moreover, activity at this cluster representing theta-to-alpha (4-12Hz) activity was predictive of trial-by-trial variations in VAS outcome ratings, i.e. lower theta-to-alpha (4-12Hz) activity was predictive of a higher treatment success.

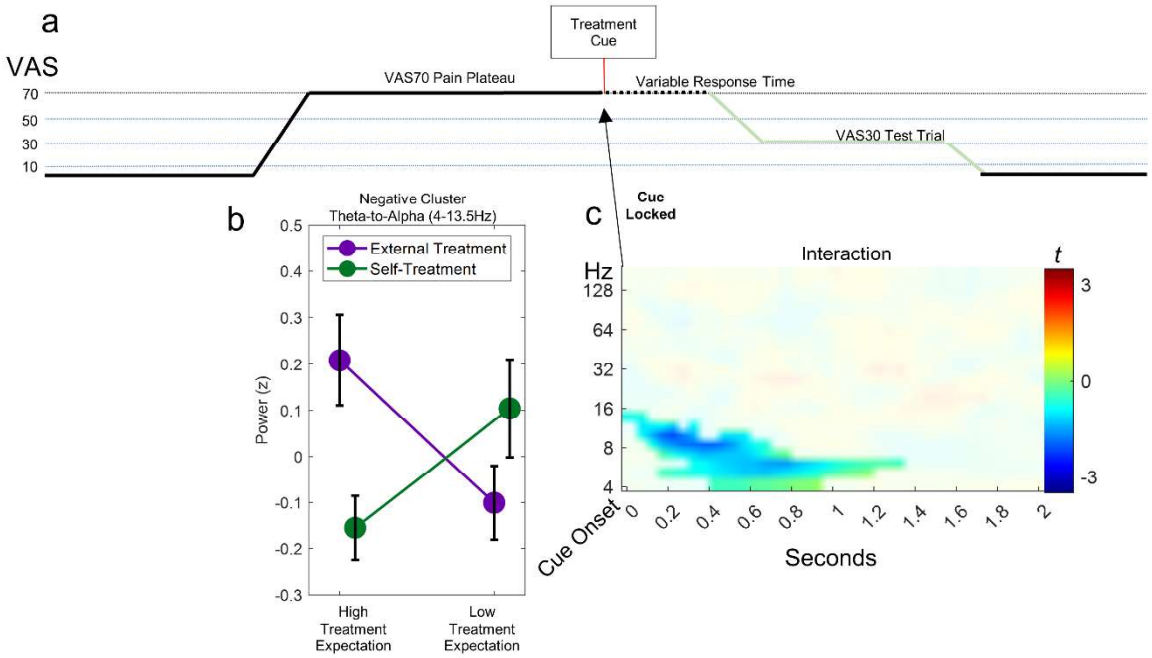



**Figure 8.** (a) Pain treatment paradigm, (b) a line plot showing EEG power averaged over the significant agency x expectation interaction cluster for self-treatment (green) and external treatment (purple), showing a significant crossed interaction, and (c) a time-frequency plot representing significant clusters of activity associated with the cue-locked interaction (agency x expectation) cluster. Time-frequency plots are averaged *t*-values over all channels including significant data points of the respective clusters of activity. Not significant data points are masked out. Colors represent *t*-values.

## Discussion

In the present study, we demonstrated greater pain relief of self-treatment and high treatment expectations, but no interaction between these factors. Bayesian model

selection provided strong evidence for a shift model (shift in the mean of the likelihood term or a shift in the mean of the prior term by self-treatment) over the likelihood precision modulation model (modulation of precision of sensory consequences of self-generated outcomes). These effects also manifested in EEG data: differences in stimulus intensity in VAS10 versus VAS50 conditioning trials were associated with differences in the theta (4-8Hz), alpha-to-beta (8-30Hz) and low gamma bands (30-50Hz). Agency modulated low frequency oscillatory responses in the alpha-to-beta range (8-30Hz) and low gamma (30-50Hz) responses at treatment outcome.

We took great care to match both conditions (self- versus external treatment) with respect to cognitive and motor demands and adjusted the trials for visual, cognitive, and motor components. We designed this experiment in a way that both conditions afforded a motor response by the participant, i.e. in self-treatment, the treatment was started by the participant by a button press, whereas in external treatment, the participants had to acknowledge the external-treatment trial by a button press. In addition, both buttons (treatment and acknowledge buttons) were displayed on the screen. Hence, during both self-treatment and external treatment, a single button press was made by the subject - either to start the treatment or to acknowledge the (supposed) experimenter-initiated treatment. In addition, the correspondingly pressed button (by the experimenter) lit up on the screen, i.e. in both conditions two buttons changed to green, as in external treatment, self-treatment was "acknowledged" by the experimenter. By doing so, we created precise temporal contingencies between the button press initiating self-treatment and the decrease of temperature. In the external treatment condition, the participant's button press to acknowledge external treatment was not locked to the decrease in temperature, instead the change of the displayed button to green by the "start" given by the experimenter was in precise contingency with the stimulus decrease. This

541 represents a typical ingredient of control or agency, namely the temporal contingency of
542 an outcome with respect to an action. Consequently, as the motor component was
543 precisely locked to the treatment in self-treatment but not in the external treatment
544 condition, this difference could influence the pain experience.

545 Concerning the role of agency, we expected an increased treatment efficacy when
546 treatment was self-generated[8,16]. Indeed, data from both experiments support this
547 hypothesis: self-treatment was associated with lower post-treatment VAS ratings as
548 compared to external treatment, despite identical objective stimulus intensity. Based on
549 previous studies, we also expected that pain perception was modulated by
550 expectation[32,34,37–40]. Both experiments support this hypothesis and showed a graded
551 effect of expectation, i.e. high treatment expectations were associated with a higher
552 treatment success than low treatment expectations.

553 Crucially, our experiment was designed to investigate the mechanism underlying
554 improved treatment efficacy when treatment was self-initiated. In a Bayesian sense, pain
555 perception can be seen as the integration of expectation (prior) and nociceptive input
556 (likelihood), with the precision of each term determining the amount of its contribution.
557 Here, we investigated on which of these parameters agency acted. This has been
558 theoretically motivated: The forward model[30] posits that small prediction errors during
559 self-generated movement lead to a percept of a less intense sensation, relative to
560 externally generated unpredicted outcomes; applied to our pain protocol, this suggests an
561 improvement in treatment outcome by self-treatment from which we theoretically
562 derived a (likelihood) mean shift model (see Fig. 1d). Alternatively, it has been suggested,
563 in the context of active inference, that action necessarily entails a decrease in precision of
564 self-generated sensory consequences[31]. Thus, because of the reduced sensory nociceptive

(likelihood) precision in the likelihood precision modulation model, the effect of expectation should increase. Under these conditions, we would expect that the relative influence of prior expectations would be increased relative to the sensory evidence, which would be attenuated in precision, as represented in our likelihood precision modulation model. Therefore, self-treatment should lead to a greater influence of treatment expectation relative to external treatment[24]. From a statistical perspective, this would manifest as an interaction between agency and expectation, i.e. larger differences between low and high treatment expectations in self-treatment as compared to external treatment (see Fig. 1e).

Our data showed clear effects of sensory attenuation and treatment expectations in two experiments with different pain rating modalities, and no significant interaction between sensory attenuation and treatment expectation effects. This clearly favors the mean shift model for self-initiated pain treatment. This was formally assessed by Bayesian model comparisons, which strongly favored the mean shift model over the likelihood precision modulation model and all other models. This is in agreement with a study by Woo et al. (2017) where choices regarding pain stimulation showed additive reductions of pain by control and expectations[41], which can also be interpreted as a prior or likelihood shift by control in a Bayesian sense.

The Bayesian likelihood (mean) shift model included a free parameter ($p_{shift}$) able to shift the mean of the likelihood distribution. However, it should be noted that the shift model can be implemented in an alternative manner, namely using a free parameter that can vary the mean of the prior distributions instead of the mean of the likelihood, which cannot be differentiated on a computational modelling basis using the Bayesian integration model. However, the EEG data with its temporal resolution can help to

589 disentangle this ambiguity. The EEG data showed a cue-locked negative cluster of activity
590 in the theta-to-alpha range (4-12Hz), which was correlated with trial-by-trial VAS ratings
591 in test trials. The early onset of this cluster (associated with cue onset) favors a
592 modulation (i.e. shift) of expectations as opposed to somatosensation and provides
593 evidence that the modulation by agency affects the prior rather than the likelihood term
594 in this model.

595 EEG data analysis was focused on two phases. Firstly, we investigated EEG power shortly
596 after cue presentation, and secondly, we explored a time frame including the pain relief
597 phase and the outcome phase, i.e. 1s before the target outcome temperature was reached
598 by the thermode and 2s after.

599 As expected for differences in physical stimulus intensities (see Ploner et al., 2017 for a
600 review), our analysis revealed a clear difference between VAS10 and VAS50 conditioning
601 trials at the treatment outcome[42]. Relative to the VAS50 condition, higher treatment
602 success (VAS10) was associated with decreased alpha-to-beta (8-30Hz) and theta (4-8Hz)
603 activity during the relief phase, while the outcome phase was associated with increased
604 alpha-to-beta and decreased theta activity for VAS10 as compared to VAS50 trials. This is
605 in line with findings associating lower alpha-to-beta power (8-30Hz) and higher theta
606 power (4-8Hz) with higher pain ratings or higher pain intensity[32,34,43–47]. Here, we extend
607 this to a treatment context, where higher alpha-to-beta power (8-30Hz) and lower theta
608 power (4-8Hz) was associated with more treatment success during the outcome phase.

609 Additionally, we observed a negative association of alpha-to-beta power and treatment
610 success during the relief phase. Importantly, using a control analysis we ruled out that
611 these clusters are associated with increased outcome latencies by ramp times leading to
612 differences in time locks in VAS10 as compared to VAS50 conditioning. In general, this

analysis revealed a very interesting pattern: While in the relief phase, decreased alpha-to-beta activity was associated with lower stimulus intensity. In the outcome phase, when the target temperature was reached, increased activity was associated with lower stimulus intensity.

A similar pattern i.e. a negative cluster starting in the relief phase in the theta-to-low gamma range (4-50Hz) and a positive cluster at the outcome phase in the alpha-to-beta range (8-30Hz) emerged when comparing self- versus external treatment trials. In the treatment phase, it is possible that latencies in reaction times for self-treatment can cause different pain plateau durations (as compared to external treatment), which can lead to unspecific time effects such as habituation. To address this concern we explicitly accounted for reaction times in our statistical model, which revealed that the cluster representing a negative association of EEG power of the agency contrast was in part explained by differences in reaction times and thus different pain plateau times as compared to external treatment. This could explain the similar negative association in self-treatment by increased habituation resulting from a fatigue of peripheral nociceptive neurons. Pain intensity would already be decreased in self-treatment as compared to external treatment in the relief phase by increased habituation[48–52], similar to an actual difference in stimulus intensity. Importantly, the cluster representing a positive association with agency was not affected by differences in reaction times. Even though behavioral ratings indicated a comparable influence of treatment expectations, we did not find expectations associated with EEG activity at the treatment outcome. Instead, theta and alpha oscillations at cue onset were differentially representing agency and expectations and were predictive of VAS outcome ratings.

We observed modulations by stimulus intensity and agency, but not by expectations in outcome-locked EEG data. This is in line with previous studies which revealed cue-related expectation effects in the alpha-to-beta band before painful stimulation[32,34], but not during painful stimulation. In another study, pain-induced alpha and gamma responses were significantly influenced by stimulus intensity but not by placebo hypoalgesia[33]. However, it has been demonstrated that expectation-based pain modulation can influence event-related pain potentials[33,40,53–55]. Overall, this suggests that expectations are associated with cue-locked effects. Expectations might be encoded in oscillatory processes of brain areas typically associated with contextual influences of top-down processing. This is in contrast to agency, which modulated activity at both, cue-locked (as an interaction) and outcome locked activity (as a main effect), suggesting influences on expectations and sensory processing.

A possible limitation of our modeling approach is that it considers the entire session and cannot reveal the dynamics e.g. of Bayesian updating across trials as can be revealed using a dynamical systems approach[56]. However, in our study, in which expectations were relatively stable throughout the experiment, temporal variations of expectations would be rather small. Since the temporal evolution of pain-related expectancies are an important factor in e.g. chronic pain conditions[25,57], future studies with amended protocols could consider these dynamics.

In conclusion, pain treatment is additively enhanced by agency and positive expectations. Sensory attenuation and objectively different stimulus intensities modify oscillatory activity at the relief and outcome phase of pain treatment, whereas expectation effects (interacting with agency) were associated with EEG activity directly following the cue. Using Bayesian model comparisons our data revealed no evidence for a decrease of

precision in self-treatment, thus favoring a mean shift as the mechanism underlying the positive effect of self-treatment.

### Experimental model and subject details

This is a human study. Information related to the subjects (e.g. sample size and sex) can be found in the Method details section of the STAR Methods of this paper.

## Lead Contact

Further information and requests for resources should be directed to the lead contact Andreas Strube (a.strube@uke.de) or to Christian Büchel (buechel@uke.de).

## Data and code availability

Data and code for this study are available on https://osf.io/q8tgj/.

Strube A., Horing B., Rose M., Büchel C., (2022) Open Science Framework ID q8tgj. Placebo and Sensory Attenuation in Pain Treatment.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## Materials availability

This study did not generate new unique materials or reagents.

### Method details

We conducted two experiments in which positive and negative treatment expectations as well as self- and external treatment were combined. In experiment 1, subjects were continuously rating their pain experience during painful stimulation and after self- or external treatment of pain. In experiment 2 with EEG recordings, we restricted the

paradigm to include only two rating phases instead of a continuous rating to avoid excessive movement.

## Subjects

In experiment 1, 29 healthy participants were enrolled. All participants gave informed consent and were paid as compensation for their participation. Applicants were excluded if one of the following exclusion criteria applied: neurological, psychiatric, dermatological diseases, pain conditions, current medication, or substance abuse. All volunteers gave their informed consent. The study was approved by the Ethics Board of the Hamburg Medical Association. Four participants had to be excluded due to adverse reactions to the capsaicin cream, leaving a final sample of 25 participants (mean age 29.3, range 19–61 years, sex: 14 female / 11 male).

The required sample size of experiment 2 was determined according to a power calculation[58] (G*Power V 3.1.9.4) based on the behavioral sensory attenuation and expectation effects in experiment 1. For the sensory attenuation effect, we observed an effect size of Cohen's $f = 0.508$ ($\eta_p^2 = 0.205$) and an effect size of $f = 0.669$ ($\eta_p^2 = 0.309$) for the expectation effect. Using a power of (1-beta) of 0.8 and an alpha level of 0.05 and assuming low correlations (0.2) among repeated measures, this leads to a required sample size of 15, taking into account the weaker agency effect. However, given the different rating in the second experiment we increased the planned number of participants to 60. Assuming the same proportion of excluded participants as in experiment 1, this allowed us to potentially detect a medium effect size[59] of $f = 0.25$ with a sample size of 53.

We enrolled 60 healthy participants in experiment 2. Five participants had to be excluded due to adverse reactions to the capsaicin cream and 1 participant had to be excluded due to technical errors during recording, leaving a final sample size of 54 (mean 28.2, range 20–60 years, sex: 34 female / 20 male).

## Thermal stimulation and capsaicin application

Both experiments started with the same preparation procedure with the application of a capsaicin cream (ABC Heat Cream, Beiersdorf AG, Germany, 750µg capsaicin/g) to the left radial forearm. Two skin patches of the size of the thermal stimulator probe were covered with the capsaicin cream for a total of 15 minutes. Thermal stimulation was performed using a 30 × 30 $mm^2$ Peltier thermode (Pathway model ATS, Medoc, Israel). The baseline temperature was set to 30°C for experiment 1 and the rise rate was set at 8°C/s for both experiments. The baseline temperature for experiment 2 was set at a lower temperature of 28°C to minimize skin irritation and attrition. After 2 blocks (experiment 1) or after the first experimental block (experiment 2), the capsaicin cream was reapplied for 5 minutes, and the stimulated skin patch was changed to avoid sensitization. In a first step, a single thermal stimulation with a slowly increasing ramp was used to familiarize the participant with the thermal stimulation. To test if the capsaicin cream was effectively reducing the pain threshold participants were asked to report the moment they felt a sensation of pain. If participants reported pain only above 46°C, the cream was reapplied for another 5 minutes on the skin patch and the initial slowly ramping heat stimulus was repeated (this applied to 2 participants in experiment 1 and to 4 participants in experiment 2).

## TENS cover story as a treatment situation

Afterwards, TENS (Transcutaneous Electric Nerve Stimulation) was established as a cover story for the treatment situation. TENS was presented as a nerve stimulation to effectively reduce pain by modulation of the nerve transmission. Putative TENS has been used to reliably generate treatment expectations in pain paradigms[27,60–62]. We provided volunteers with a deceptive "sham brochure" explaining that different stimulation frequencies result in different treatment efficacies. An electrode was attached to the elbow which was connected to an electrical current stimulator (Digitimer Ltd., model DS7A, United Kingdom). Participants were told that the electrical current stimulator needed to be individually calibrated. For this, we applied short trains of electrical currents with increasing intensity and asked the participant to report if there was a sensation; this was intended to establish the belief that the device is actually active and capable of producing said currents. Afterwards and without knowledge of the participant, the electrical current stimulator was turned off and the participants were told that the settings for optimal stimulation were found. During the experiment, no actual electrical stimulation was applied. Additionally, and to reinforce the TENS cover story, participants were asked to report if they felt a stimulation during the experiment and they were told that if this was the case, the stimulator needed to be recalibrated.

## Pain calibration

We individually calibrated the heat stimulation using an adaptive procedure to the levels of 10, 30, 50 and 70 on a Visual Analogue Scale (VAS) from -100 to 100 where a VAS of 0 represented the pain threshold. The VAS was presented on a computer screen and ratings were given using the cursor keys on a conventional keyboard. During the information

procedure at the beginning of the experiment, participants got acquainted with the VAS scale and printed screenshots of VAS scales were shown and explained. At first, participants were stimulated with 34°C, 34.5°C, 35°C and 35.5°C and were asked to report if any of these stimuli were painful. Note that temperatures required to generate pain on capsaicin sensitized skin are regularly in this temperature range. If the participant reported that the stimulation was painful, the procedure was continued with a starting temperature of 35°C, otherwise the starting temperature was set to 36°C for a stepwise procedure to find the pain threshold. Stimulus duration was set to 8s, according to the duration of VAS70 pain during the experiment. For the stepwise stimulus determination, 8 stimuli were presented with fixed reductions and increases in temperature relative to the pain threshold and participants were asked to rate the stimuli on a scale which was labeled as “normal sensation” at VAS -100, “minimally painful” at VAS 0 and “extremely painful” at VAS 100. Individual VAS levels of 10, 30, 50 and 70 were estimated using a linear regression of the VAS ratings recorded during this calibration phase. See Supplementary Data for calibration data.

## Trial design and block structure: experiment 1

Each trial was structured in 3 phases: Trial start, pain phase and treatment phase (see Fig. 2b). At trial start, an empty bar was presented in the center of the screen. The thermode temperature remained at the baseline of 30°C for 5s. Afterwards, the pain phase started, which was signaled by a filled red bar in the center of the screen. Thermode temperature was increased with a rate of 8°C/s to the temperature corresponding to the calibrated pain value of VAS70. The pain phase lasted for randomly jittered 8-10s. At the beginning of the treatment phase, a cue was presented which indicated whether high or low treatment effectivity was to be expected and whether self-treatment or external

771 treatment would occur. The cue was designed as a reduction of the centered red bar (i.e.
772 more reduction by 2/3 of the total height with high treatment expectation as compared
773 to a reduction by 1/3 of the total height with low treatment expectation) and the word
774 YOU (i.e. self-treatment) or HE (i.e. experimenter-induced, external treatment) written
775 inside the bar, indicating self-or external treatment. After a lag of 2s, 2 treatment buttons
776 were activated and appeared on the display, changing to green when pressed either by
777 the subject or automatically. The external treatment was communicated as being done by
778 the experimenter to reinforce the notion of a treatment setting but was in fact computer-
779 initiated with a naturalistic "reaction time" delay. In the case of self-treatment, the
780 participant pressed a button (A) and the treatment started with a reduction of the
781 thermode temperature to the target level of VAS50, VAS30 or VAS10, depending on the
782 condition. Meanwhile, participants received the signal of a button press (B) from the
783 experimenter as an indication that the experimenter had acknowledged the self-
784 treatment. In the case of an external treatment, a button press (B) by the participant had
785 to acknowledge the external treatment. Meanwhile, participants received an indication of
786 a button press (A) from the experimenter, signaling that the treatment has been started.
787 Participants were instructed to perform this task as soon as the treatment buttons
788 appeared on the screen. Importantly, this procedure ensured identical motor output for
789 the self-treatment and the external treatment conditions. In conditioning trials, the
790 expectation of highly effective treatment resulted in a relatively more effective treatment
791 and a reduction of the pain stimulus to the individual level of VAS10, as compared to a
792 reduction of the pain stimulus to VAS50 in conditioning trials with low treatment
793 expectations. For test trials, regardless of cued treatment effectivity, the pain stimulus
794 was reduced to VAS30. The reduction of the pain stimulus was set at -8°C/s. In total, the
795 treatment phase lasted 17-19s for a total trial duration of 32s including all 3 phases.

Additionally, participants were asked to continuously rate their pain level on a scale from 0 (minimal pain) to 100 (extreme pain) during the whole trial duration. A rating scale with a starting point at VAS0 (i.e. position of a red rating indicator) appeared ranging from VAS0, labelled as “minimally painful” to VAS100, labelled as “extremely painful”. The 2 buttons used for the rating were represented on the screen and were lighting up when pressed on the keyboard. After completion of the rating phase, the heat stimulus was reduced to the baseline temperature for the remaining intertrial interval of 18s.

During experiment 1, 4 experimental blocks were presented. Each block consisted of a total of 26 trials. It started with 8 conditioning trials, 4 of which were associated with low treatment success, 4 with high treatment success, each with the respective cues. After the conditioning trials, 3 micro blocks were presented consecutively. Each micro block consisted of 6 trials of following types:

(1) 1 conditioning reinforcement trial with high expectation of treatment success with actual high treatment success (reduction of pain from VAS70 to VAS10).

(2) 1 conditioning reinforcement trial as a reinforcement with low expectation to treatment success with actual low treatment success (reduction of pain from VAS70 to VAS50).

(3) 4 test trials with medium treatment success (reduction in pain from VAS70 to VAS30). I.e.:

a. 1 trial with high expectation of treatment success and self-treatment with a reduction of pain from VAS70 to VAS30.

b. 1 trial with high expectation of treatment success and external treatment with a reduction of pain from VAS70 to VAS30.

c. 1 trial with low expectation of treatment success and self-treatment with a reduction of pain from VAS70 to VAS30.

d. 1 trial with low expectation of treatment success and an external treatment with a reduction of pain from VAS70 to VAS30.

The order of trials within these micro blocks was randomized. Randomization was constrained so that a trial was not directly followed by the same type of trial, e.g. there were no 2 consecutive low expectation self-treatment test trials. For the 8 conditioning trials at the beginning of a block, it was ensured that at most 2 consecutive conditioning trials with the same condition (e.g. high treatment success) occurred.

In total, 4 experimental blocks were presented. During the first block conditioning and reinforcement trials were either exclusively self-treatment trials or external treatment trials. This was switched after 2 blocks, i.e. if the first 2 blocks were self-conditioning blocks the last 2 blocks were external conditioning blocks and vice versa.

Before the first experimental block was presented, 4 training trials were performed, during which the illusion of treatment was demonstrated by pressing the button connected to the heat stimulation device for pain reduction. At this stage, an actual button press by the experimenter was required during external training trials to establish the illusion of a direct link between TENS and the button press as for the participant. To do so, the experimenter sat next to the participant and demonstratively pressed the required button on the keyboard of the participant. In the remainder of the experiment, external "button presses" were performed by the computer unbeknownst to the subject.

## Trial design and block structure: experiment 2

During experiment 2 (see Fig. 2c), the paradigm was split into 5 phases: Trial start (4s), pain phase (8s), pain rating phase (6s), treatment phase (8s) and treatment rating phase (6s). Rating scales and related rating buttons on the screen were only presented during rating phases. During the pain rating phase, a red indicator on the VAS was presented with a random starting position. During the treatment rating phase, the final pain rating position of that red indicator was presented for orientation alongside a new green indicator initially appearing at a random position. The green indicator was used to rate the treatment outcome. In the treatment phase of experiment 2, treatment buttons were presented and activated simultaneously with the treatment cue without a jittered lag (as compared to experiment 1).

In total, 2 experimental blocks were presented. The first block consisted of conditioning and reinforcement trials which were either exclusively self-treatment trials or external treatment trials. This was switched after one block, i.e. if the first block was a self-conditioning block the second block was an external conditioning block, and vice versa. In total, 56 trials were presented per block, consisting of 8 conditioning trials followed by 8 micro blocks each containing both reinforcers (VAS10 and VAS50 conditioning trials) and each of the 4 test trial types (self- versus external conditioning, low versus high treatment expectation). Trials were presented with an intertrial interval of 4s. The first experimental block was presented after 4 training trials which were performed as in experiment 1.

## Questionnaire Data

After experiment 2 was concluded, participants were asked to complete several questionnaires. We included the BDI-V (simplification of the Beck Depression Inventory), LSHS-R (Launay-Slade Hallucination Scale – German revised version), STAI-X1 (State) and STAI-X2 (Trait; State-Trait-Anxiety-Scale), FKK (German locus of control scale), SWE (German General Self-Efficacy Scale) and PCS (Pain Catastrophizing Scale – German translation) scales[63–71]. Pearson product-moment correlation coefficients were computed to assess the relationship between questionnaire data and agency and placebo benefits. Agency benefits were defined as the difference between post-treatment VAS ratings of self-treatment and external treatment conditions. Placebo benefits were defined as the difference between post-treatment VAS ratings of high treatment expectation and low treatment expectation conditions. See Supplementary Table 1 for a summary of the correlational results of the questionnaire data.

## EEG recording

EEG data were acquired using a 64-channel Ag/AgCl active electrode system (ActiCap64; Brain Products GmbH, Germany) placed according to the extended 10–20 system[72]. 60 electrodes were used of the most central scalp positions. The EEG was sampled at 500 Hz, referenced at FCz, and grounded at Iz. For removal of ocular movement artifacts, horizontal and vertical bipolar electrooculogram (EOG) were recorded using the 4 remaining electrodes.

## EEG preprocessing

The data analysis was performed using the Fieldtrip toolbox for EEG/MEG analysis[73]. For preprocessing, data were epoched and time-locked to the onset of the cue signaling the start of the treatment phase. Each epoch was centered (subtraction of the temporal mean) and included a time range of 19s before and 9s after trigger onset (starting with the empty cue signaling the start of the trial up to the end of the treatment phase).

We employed a preprocessing approach by Hipp et al. (2002)[74] by splitting the data into 2 band-pass filtered sub-sets from 4 to 34Hz for low frequencies and from 16 to 250Hz for high frequencies. This enabled efficient separation of low- and high frequency artifacts in subsequent ICA analysis. EEG epochs were visually inspected, and trials contaminated by artifacts due to gross movements or large technical artifacts were removed. Trials contaminated by eye-blinks, muscle activity, technical artifacts or movements were corrected using an independent component analysis (ICA) algorithm[75,76] after careful inspection of topographies, power spectra and relation of ICA time courses to the temporal structure of the experiment. Artifactual components were removed before the remaining components were back-projected and resulted in corrected data. Subsequently, the data were re-referenced to a common average of all EEG channels and the previous reference channel FCz was reused as a data channel. Finally, epochs were visually screened and trials with remaining artifacts were excluded from analysis.

Before time–frequency transformations for data analysis were performed on the cleaned datasets, the time axis of single trials was shifted to create cue-locked and outcome-locked data. For cue-locked data, we set the onset of the cue signaling to the start of the treatment phase as $t = 0$. Outcome-locked data takes individual differences in response time into account and sets $t = 0$ to the time point when the thermode reached the treatment target

temperature (calibrated VAS10, VAS30 and VAS50 levels for low conditioning, test and high conditioning trials, respectively). Trials were excluded if this duration (from cue onset to treatment outcome) was longer than 6s. This allowed us to create an analysis window of 2s in subsequent time-frequency analysis without contamination by the subsequent rating phase.

## EEG spectral analysis

Spectral analysis was adapted from Hipp and colleagues[74]. This approach ensured a homogenous sampling and smoothing in time and frequency space. We calculated spectral estimates for 23 logarithmically scaled frequencies ranging from 4 to 181 Hz (0.25 octave increments) for the pain phase and treatment phase in 0.05s steps. For cue-locked data, this included the treatment phase from cue onset up to 2s after cue onset. For treatment-locked data, this included the relief phase from 1s before the treatment outcome (target temperatures of VAS10, VAS30 or VAS50) was reached, and the outcome phase up to 2s after the target temperature was reached. Using the multitaper (DPSS) approach, we set the temporal and spectral smoothing to match 250ms and 3/4 octave, respectively. For frequencies below 16 Hz, we employed 250ms temporal windows and varied the number of Slepian tapers to approximate a 3/4 octave spectrum smoothing. We changed the time window for frequencies below 16 Hz to achieve a frequency smoothing of 3/4 octaves with a single taper. We computed the frequency transform using high- and low-frequency data for frequencies above and below 25 Hz, respectively. Analysis was then continued with the combined spectral data after averaging of spectral estimates per block and condition over trials for each subject.

For the baseline correction of time–frequency data, the mean and standard deviation were estimated (for each subject/channel/frequency combination, separately) from 0.5 to 7.5s of the pain phase (i.e. increases and decreases in EEG power activity indicate deviations from EEG power during painful stimulation). The mean spectral estimate of the baseline was then subtracted from each data point, and the resulting baseline-centered values were divided by the baseline standard deviation (classical baseline normalization – additive model[77]).

### Quantification and statistical analysis

All statistical parameters described in this section are reported in the Results section or in the Supplementary Data.

## Behavioral data analysis

For experiment 1, we performed analysis on the continuous VAS rating by simply using the last data point of each trial. For experiment 2, we performed analysis on the single post-treatment VAS rating. Here, we have 2x2 conditions for test trials (low versus high treatment expectation / self- versus external treatment) and 2 conditions for conditioning trials. Firstly, we conducted a 3x1 repeated measures ANOVA with post-hoc t-tests to evaluate the differences between VAS10 (conditioning), VAS30 (test) and VAS50 (conditioning) conditions, respectively. Secondly, we conducted a 2x2 repeated measures ANOVA to evaluate differences between the different test conditions (low versus high expectation / self- versus external treatment) and the interaction term of expectation and agency.

## Bayesian integration models of placebo pain treatment

For model-based analysis of our post-treatment VAS rating data, we designed Bayesian integration models of pain perception in placebo pain treatment (see Büchel et al., 2014 for a review) in accordance with the likelihood shift model and the likelihood prior modulation model[24]. In the Bayesian formulation of pain perception, Bayes' theorem is used to estimate the level of perceived pain, taking precision-weighted prior experiences into account (Fig. 1a and Eq.1). Formally, the model integrates a prior with a likelihood to estimate a posterior. Both the prior and the likelihood were approximated by normal distributions allowing for an analytical integration using normal-normal conjugate priors to estimate the normal posterior.

(Eq.1)

$$\mu_{posterior} = \frac{\mu_{prior} * \sigma^2{}_{likelihood} + \mu_{likelihood} * \sigma^2{}_{prior}}{\sigma^2{}_{likelihood} + \sigma^2{}_{prior}}$$

$$\sigma^2{}_{posterior} = \frac{\sigma^2{}_{likelihood} * \sigma^2{}_{prior}}{\sigma^2{}_{likelihood} + \sigma^2{}_{prior}}$$

With respect to the behavioral data, our model predicted the painfulness of the test phase post-treatment VAS ratings (posterior) by integrating conditioning post-treatment VAS ratings as a prior (mean and variance derived from VAS10 and VAS50 post-treatment VAS ratings for high and low treatment expectation conditions, respectively) with an individual estimate of the likelihood (average of VAS10 and VAS50 parameters for each subject). Gaussian approximation of the rating data was performed by fitting a Gaussian cumulative probability density functions to the cumulative sum of the ratings using a robust grid search[27].

For the estimation of the posterior parameters in self-treatment trials we created two derived models, based on a shift of the likelihood and a modulation of the likelihood precision. For the likelihood shift model, we included a free parameter $p_{shift}$ to enable a likelihood shift (Eq.2; Fig. 1b):

(Eq.2)

$$\mu_{posterior} = \frac{\mu_{prior} * \sigma^2{}_{likelihood} + (\mu_{likelihood} + p_{shift}) * \sigma^2{}_{prior}}{\sigma^2{}_{likelihood} + \sigma^2{}_{prior}}$$

$$\sigma^2{}_{posterior} = \frac{\sigma^2{}_{likelihood} * \sigma^2{}_{prior}}{\sigma^2{}_{likelihood} + \sigma^2{}_{prior}}$$

For the likelihood precision modulation model we included a free parameter $p_{precision}$ to represent a modulation of likelihood variance by self-treatment. Under this model, posterior parameters are estimated by the following equations (see also Fig. 1c):

(Eq.3)

$$\mu_{posterior} = \frac{\mu_{prior} * (\sigma^2{}_{likelihood} * p_{precision}) + \mu_{likelihood} * \sigma^2{}_{prior}}{(\sigma^2{}_{likelihood} * p_{precision}) + \sigma^2{}_{prior}}$$

$$\sigma^2{}_{posterior} = \frac{(\sigma^2{}_{likelihood} * p_{precision}) * \sigma^2{}_{prior}}{(\sigma^2{}_{likelihood} * p_{precision}) + \sigma^2{}_{prior}}$$

We used a variational Bayesian inference to estimate the parameters of all models using the VBA toolbox[78] for Matlab (R2021a). Note that we also tested a prior precision modulation model where the free parameter $p_{precision}$ was modulating prior precision. We used uninformative priors for both parameters ($p_{precision} \sim Normal(1,1000)$ and $p_{shift} \sim Normal(0, 1000)$. In addition, we fitted a full combined model of all three free parameters (i.e. $p_{shift}$, $p_{precision}$ at prior, $p_{precision}$ at likelihood), a combined model including the $p_{shift}$ and

$p_{precision}$ at prior parameters, a combined model including the $p_{shift}$ and $p_{precision}$ at likelihood parameters, and a null model (Eq.1) in which all parameters were "constrained" through their priors ($p_{precision} \sim Normal(1, 1e\text{-}20)$ and $p_{shift} \sim Normal(0, 1e\text{-}20)$). Given our behavioral post-treatment VAS rating data (i.e. empirical posterior), VBA recovers an approximation to both the posterior density on unknown variables ($p_{precision}$ and $p_{shift}$ for the likelihood precision modulation model and the likelihood shift model, respectively) and the log model evidence (which is used for model comparison). We used a random effects (RFX) Bayesian model selection approach[35,36] to estimate the overall posterior model probability across subjects. Finally, we estimated the protected exceedance probability as a metric for the Bayesian model comparison of all candidate models[35].

## EEG data analysis

Here, we analyzed the effects of 2 phases of pain treatment. Firstly, we wanted to analyze effects associated with the treatment cue indicating low or high treatment success and self- or external treatment. For this analysis, we used a cue-locked analysis window of 2s after the onset of the cue. Secondly, we wanted to evaluate the relief phase and treatment outcome based on low or high treatment expectations, agency, and their interaction. As the treatment outcome occurred at highly variable time points based on the response, we analyzed –1 to 2s in relation to the time point when the thermode reached the treatment target temperature (where -1 to 0s was defined as the relief phase and 0 to 2s was defined as the treatment outcome phase).

We corrected all statistical tests in electrode space for multiple comparisons using non-parametrical permutation tests of clusters[79]. Samples (exceeding the threshold of $p < .05$)

were clustered in connected sets on the basis of temporal (i.e. adjacent time points), spatial (i.e. neighboring electrodes), and spectral adjacency. Clustering was restricted in a way that only samples were included in a cluster which had at least 1 significant neighbor in electrode space (i.e. at least one neighboring channel also had to exceed the threshold for a sample to be included in the cluster). Neighbors were defined by a template provided by the Fieldtrip toolbox corresponding to the used EEG montage.

A cluster value was defined as the sum of all statistical values of included samples. Monte Carlo sampling was used to generate 1000 random permutations of the design matrix by shuffling of condition labels per subject, and statistical tests were repeated in time–frequency space with the random design matrices. The probability of a cluster from the original design matrix (*p*-value) was calculated by the proportion of random design matrices producing a cluster with a cluster value exceeding the original cluster. Muscular and ocular electrodes were excluded from the cluster analysis.

Further, we wanted to explore correlations of between-subject time-frequency responses and benefits of high treatment expectation versus low treatment expectation in post-treatment VAS ratings, as well as sensory attenuation model parameters. For each participant, we calculated the placebo benefit by the within-subject difference of *z*-normalized post-treatment VAS ratings between high and low treatment expectations in test trials. For the benefit of agency, we used the single subject mean estimate of the $p_{shift}$ parameter from the VBA model inversion procedure, accordingly.

Any positive or negative cluster in correlation analysis of EEG power during test trials would indicate an association with placebo or agency benefits. Here, a p-value of $p < 0.05$ obtained from the Pearson's correlation test statistic as implemented in the Fieldtrip toolbox was used as a threshold for clustering.

## Linear mixed effects models of EEG clusters

To better quantify the properties of EEG clusters (see Results), we assessed if these clusters were associated with trial-by-trial pain reports. In this analysis, we focused on VAS30 test trials which give the advantage that they are not confounded by different temperatures. I.e., the intensity contrast of VAS10 and VAS50 conditioning is confounded with actual differences in temperatures calibrated to VAS10 and VAS50. Considering only test trials, the temperature was kept constant, meaning that differences in EEG power are attributable to cognitive processes other than pure sensory processes by nociceptive intensity differences.

To perform a trial-by-trial analysis, we had to generate single-trial time-frequency data. We used the same parameters as in the original time-frequency analysis to generate cue-locked and outcome-locked time-frequency power estimates. It should be noted that the preprocessing in the original analysis was done on 2 separate data sets split by frequency (<=25Hz and >25Hz; see STAR Methods), so this did not necessarily mean that the same trials at both low and high frequency ranges were included for each subject. To allow a single trial analysis across these clusters (which were in both high and low frequency ranges) we only included trials for the single trial analysis that were present in both high and low frequency ranges for each subject. Overall, we removed 11.2% of trials (on average 1.51 trials per subject per test condition) in the low frequency range and 9.9% of trials (on average 1.32 trials per subject per test condition) in the high frequency range. A baseline correction was performed afterwards on each trial separately with the same basis as in the original analysis. I.e., for each condition we used the mean and standard deviation over the averaged trials for baseline correction via *z*-standardization for each single trial time-frequency power estimates, separately.

As a next step, we used these single trial time-frequency power estimates to produce a cluster value for each subject-trial-cluster combination. We extracted the cluster properties from clusters resulting from our original cluster analysis. We cumulated the power estimate of each significant datapoint (channel-time-frequency combination) from our original clusters separately for each trial and averaged this accumulated power estimate over the total amount of significant data points to create an average single-trial cluster value for each subject and trial.

Model estimation used a linear mixed-effects model as implemented in MATLAB (R2021a) in the fitlme function. In total, we included 2 clusters (see Results) in the linear mixed-effects model: The positive outcome-locked cluster of the agency contrast (self-treatment versus external treatment in test trials averaged over predictive cues) and the cue-locked interaction contrast of expectations (i.e. predictive cues) and agency. Negative outcome-locked clusters of the agency contrast were excluded based on a confound with reaction times.

Then, we fitted a linear mixed-effects model for VAS outcome ratings at VAS30 test trials, with fixed effects for each averaged single-trial cluster values, and a random effect for the intercept. A *p*-value of .05 was considered as threshold of significance.

## Reaction time and ramp time analysis

In this paradigm, pain treatment was either self-administered (i.e. self-treatment) or externally administered (external treatment) which underlie individual differences in latencies. These differences in latencies (i.e. by faster or slower responses) are associated with differences in pain plateau durations as a reduction of temperature was directly coupled to these responses. As continuous pain leads to habituation or sensitization[48–52],

EEG and behavioral effects of agency could be confounded by these processes, e.g. a longer plateau duration in the self-treatment condition might explain differences in EEG data. To test for confounding of the results, we examine the behavioral and EEG effects in a simple linear regression model. We tested agency benefits (self-treatment minus external treatment post-treatment VAS rating) averaged for each participant in a linear regression model with the average reaction time of each participant as a regressor for experiment 1 and 2. Also, we assessed if activity at different EEG clusters (see Results) were associated with latency differences.

Firstly, we assessed if clusters associated with differences between VAS10 and VAS50 conditioning were associated with outcome latency differences due to different ramp times for VAS10 and VAS50 conditioning trials. As VAS10 and VAS50 conditioning temperatures were individually calibrated, there were differences in ramp times between participants for VAS10 and VAS50 pain levels. For example, a decrease to a VAS50 temperature of 43°C from a VAS70 plateau of 47°C takes 500ms whereas a decrease to 41°C to the VAS10 level would take 750ms. Naturally, this leads to different outcome latencies based on individual ramp time differences between the VAS50 and VAS10 levels. Simple linear regression was used to test if ramp time differences significantly predicted the difference between VAS10 and VAS50 conditioning in averaged EEG power at each cluster.

As a next step, we assessed if EEG clusters resulting from the *t*-contrast of agency (self- versus external treatment) at VAS30 test trials could be explained by reaction time differences. Here, ramp time differences are not relevant as each condition tested in this contrast was associated with the same ramp time leading to VAS30 temperatures for self- and external test trials alike.

In summary, our cluster analysis revealed 2 separate clusters of activity for the agency contrast (i.e. self-treatment versus external treatment test trials) in the time-frequency EEG data at the treatment outcome (see Results). For self- and external conditions, we averaged the activity of each significant sample (each significant channel-time-frequency combination) included in the respective cluster per subject.

We tested the difference between self-treatment and external treatment of each participant in a linear regression model with the average reaction time of each participant as a regressor. Simple linear regression was used to test if reaction times significantly predicted the difference between self-treatment and external treatment in averaged EEG power of cluster samples in each frequency band.

Finally, we tested the cluster resulting from the interaction contrast (see Results) in the same manner. Here, we calculated the interaction contrast in EEG power per subject and a simple linear regression was used to test if reaction times significantly predicted the averaged power of cluster samples associated with the interaction effect. A *p*-value of .05 was considered as a threshold of significance in all tests.

## Additional resources

Experiment 2 was preregistered at the German Clinical Trial Register (ID: DRKS00025541).


## Acknowledgements

CB is supported by ERC-AdG-883892-PainPersist and DFG SFB 289 project A02. MR is supported by DFG SFB 289 project A03 and DFG SFB TR 169 project B3. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 422744262–TRR 289.

## Ethics

Human subjects: All volunteers gave their informed consent. The study was approved by the Ethics board of the Hamburg Medical Association (PV3892).

## Declaration of Interests

The authors report no conflict of interest.

## Author contributions

A.S.: Conceptualization, Data curation, Software, Formal analysis, Investigation, Visualization, Methodology, Writing - original draft, Project administration, Writing - review and editing; B.H.: Conceptualization, Software, Methodology, Writing - review and editing; M.R.: Resources, Methodology; C.B.: Conceptualization, Resources, Formal analysis, Supervision, Funding acquisition, Validation, Visualization, Methodology, Project administration, Writing - review and editing.

# Supplementary Data

## Calibration data: experiment 1

During experiment 1, pain levels were calibrated to achieve VAS10 (M = 38.1°C, SD = 3.5°C, Min = 31.8°C, Max = 44.8°C), VAS30 (M = 39°C, SD = 3.5°C, Min = 32.2°C, Max = 45.3°C), VAS50 (M = 39.9°C, SD = 3.6°C, Min = 32.5°C, Max = 46.2°C) and VAS70 (M = 40.8°C, SD = 3.8°C, Min = 32.8°C, Max = 47.2°C) pain levels (for highly effective conditioning, test trials, weakly effective conditioning and VAS70 pain stimulation, respectively).

## Calibration data: experiment 2

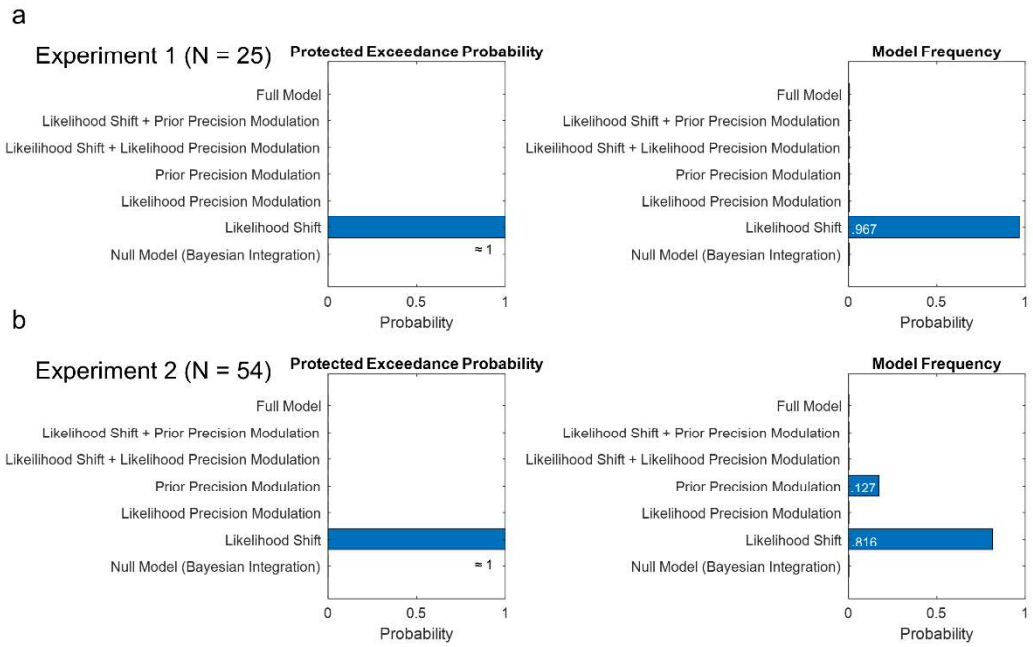
During experiment 2, pain levels were calibrated to achieve VAS10 (M = 38.2°C, SD = 3.1°C, Min = 31.72°C, Max = 44.5°C), VAS30 (M = 39°C, SD = 3.1°C, Min = 32.1°C, Max = 45.3°C), VAS50 (M = 38.19°C, SD = 3.1°C, Min = 32.5°C, Max = 46.1°C) and VAS70 (M = 40.5°C, SD = 3.2°C, Min = 32.8°C, Max = 46.9°C) pain levels (for highly effective conditioning, test trials, weakly effective conditioning and pain stimulation, respectively).

## EEG data analysis: interaction

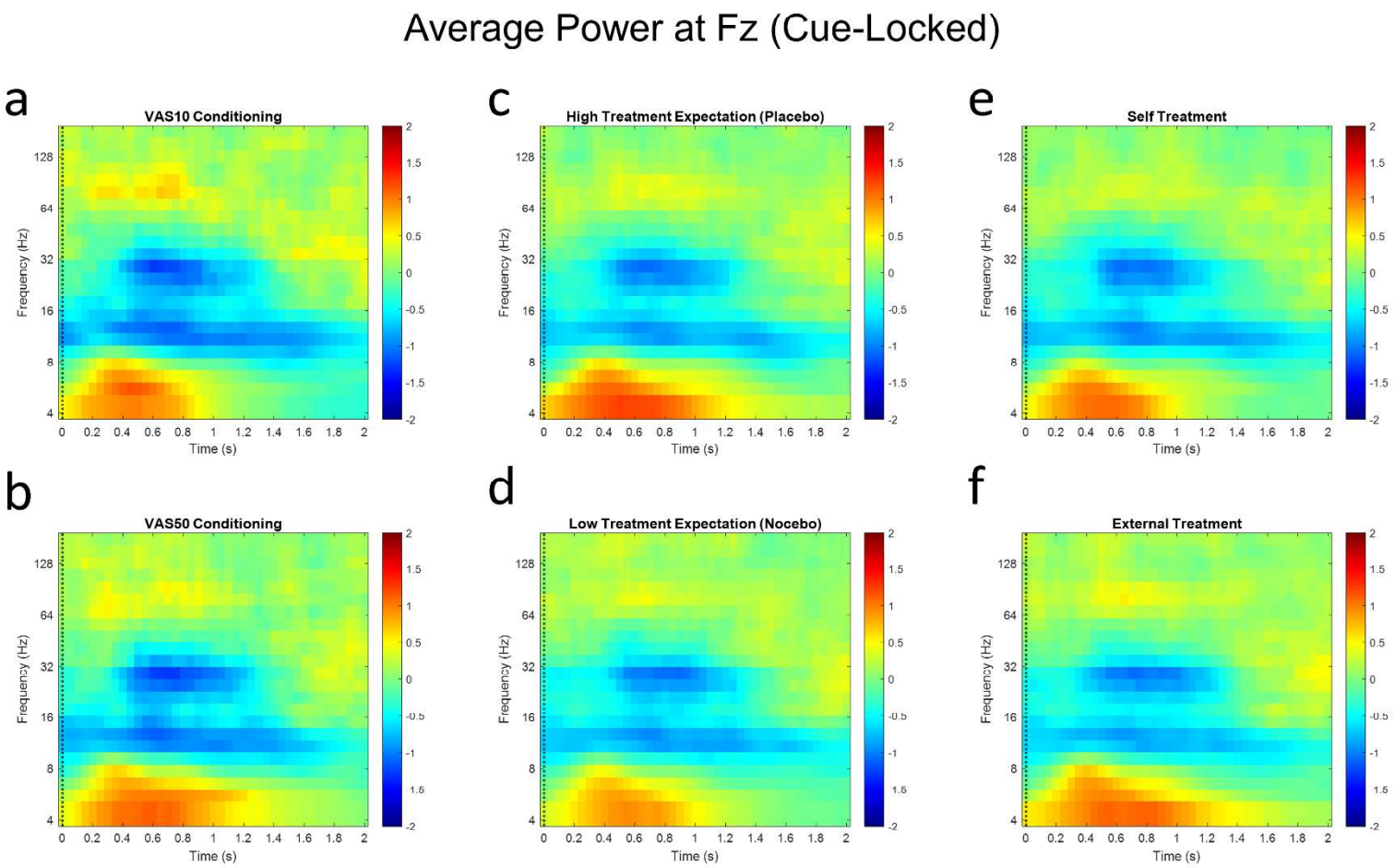
We conducted post-hoc *t*-tests to confirm the crossed interaction of agency and expectations at cue-locked EEG data. Post-hoc *t*-tests confirmed a crossed interaction where all 4 comparisons were significant, i.e. self-treatment with high treatment expectation trials were associated with lower EEG power than self-treatment trials with low treatment expectations ($t(53) = -3.76$, $p < 0.001$), whereas external treatment trials with high treatment expectations were associated with higher EEG power than external treatment trials with low treatment expectations ($t(53) = 4.86$, $p < 0.001$). Also, high treatment expectation trials with self-treatment were associated with lower EEG power

that high treatment expectation trials with external treatment ($t(53) = -6.03$, $p < 0.001$) and low treatment expectation trials with self-treatment were associated with higher EEG power than low treatment expectation trials with external treatment ($t(53) = 2.66$, $p = 0.01$).

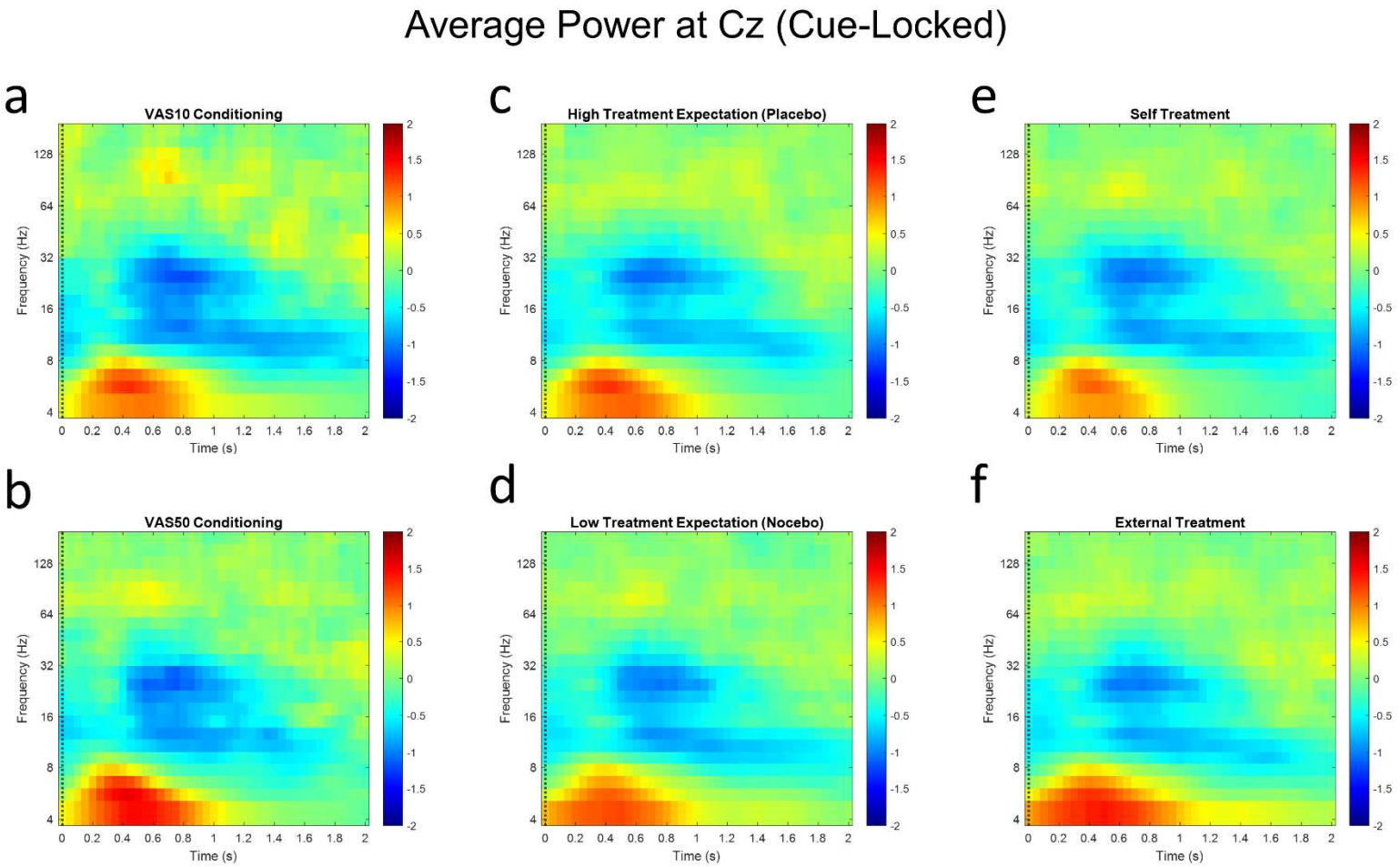
## Supplementary Figures



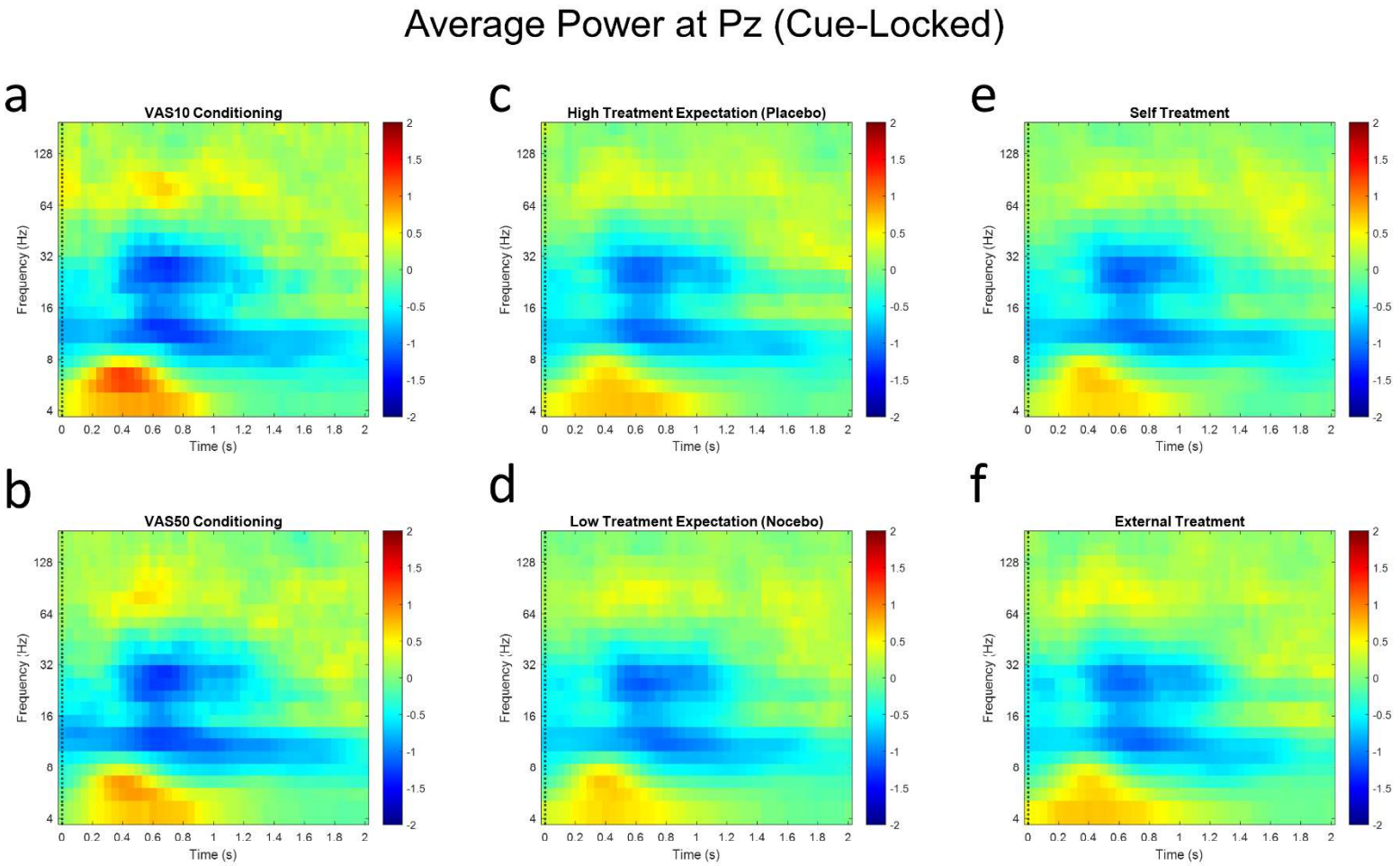

**Supplementary Figure 1.** Results of the Bayesian model comparison including all models of (a) experiment 1 (N=25) and (b) experiment 2 (N=54) showing (left) protected exceedance probabilities and (right) model frequencies.




**Supplementary Figure 2.** Time-frequency plots represent averaged, baseline-corrected, cue-locked power values for each condition at Fz. (a) VAS10 versus (b) VAS50 conditioning, (c) high versus (d) low treatment expectation and (e) self- versus (f) external treatment. Warm colors represent positive *z*-values (increase of power compared to baseline) and cold colors represent negative *z*-values (decrease of power compared to baseline).



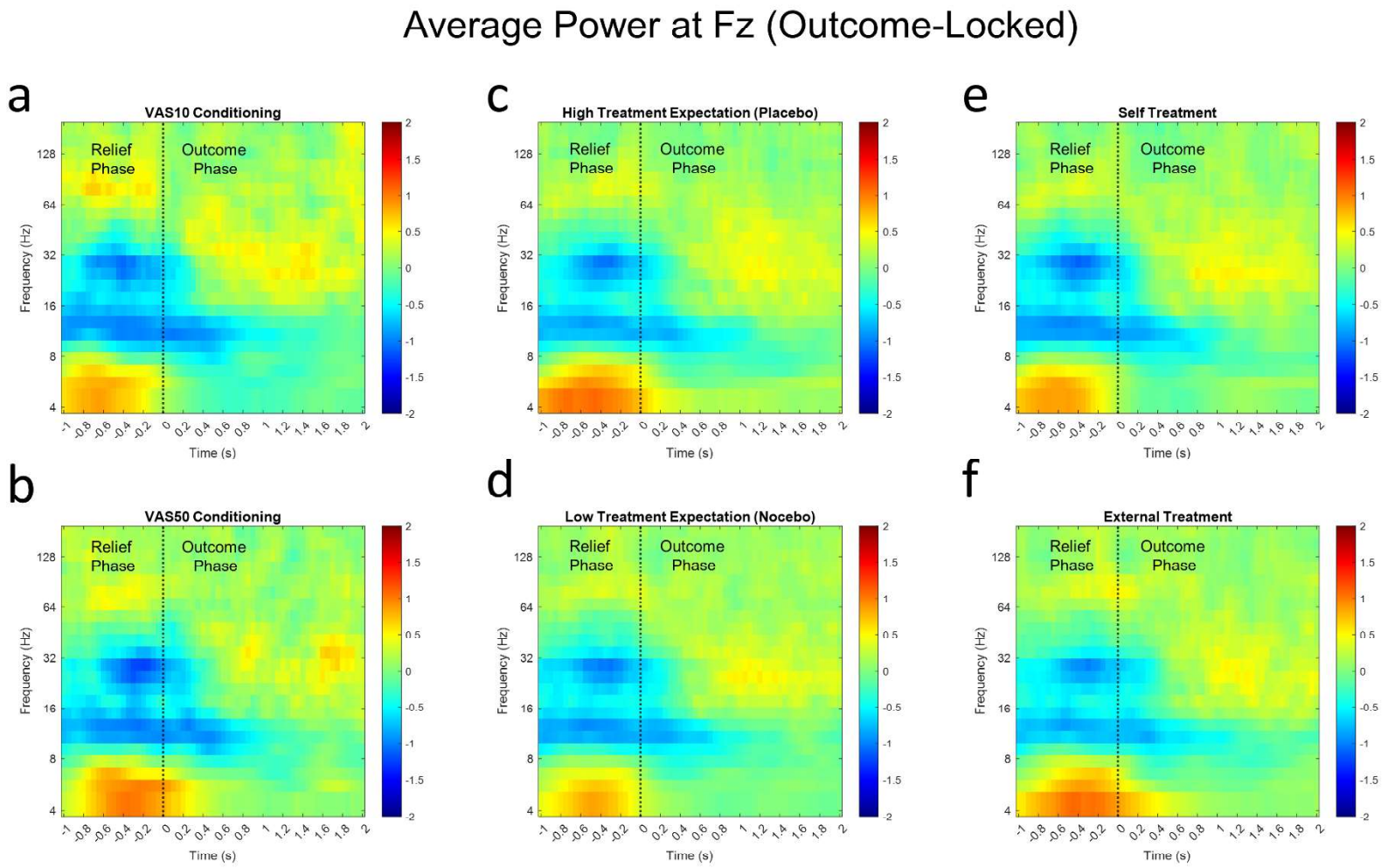

**Supplementary Figure 3.** Time-frequency plots represent averaged, baseline-corrected, cue-locked power values for each condition at Cz. (a) VAS10 versus (b) VAS50 conditioning, (c) high versus (d) low treatment expectation and (e) self- versus (f) external treatment. Warm colors represent positive *z*-values (increase of power compared to baseline) and cold colors represent negative *z*-values (decrease of power compared to baseline).



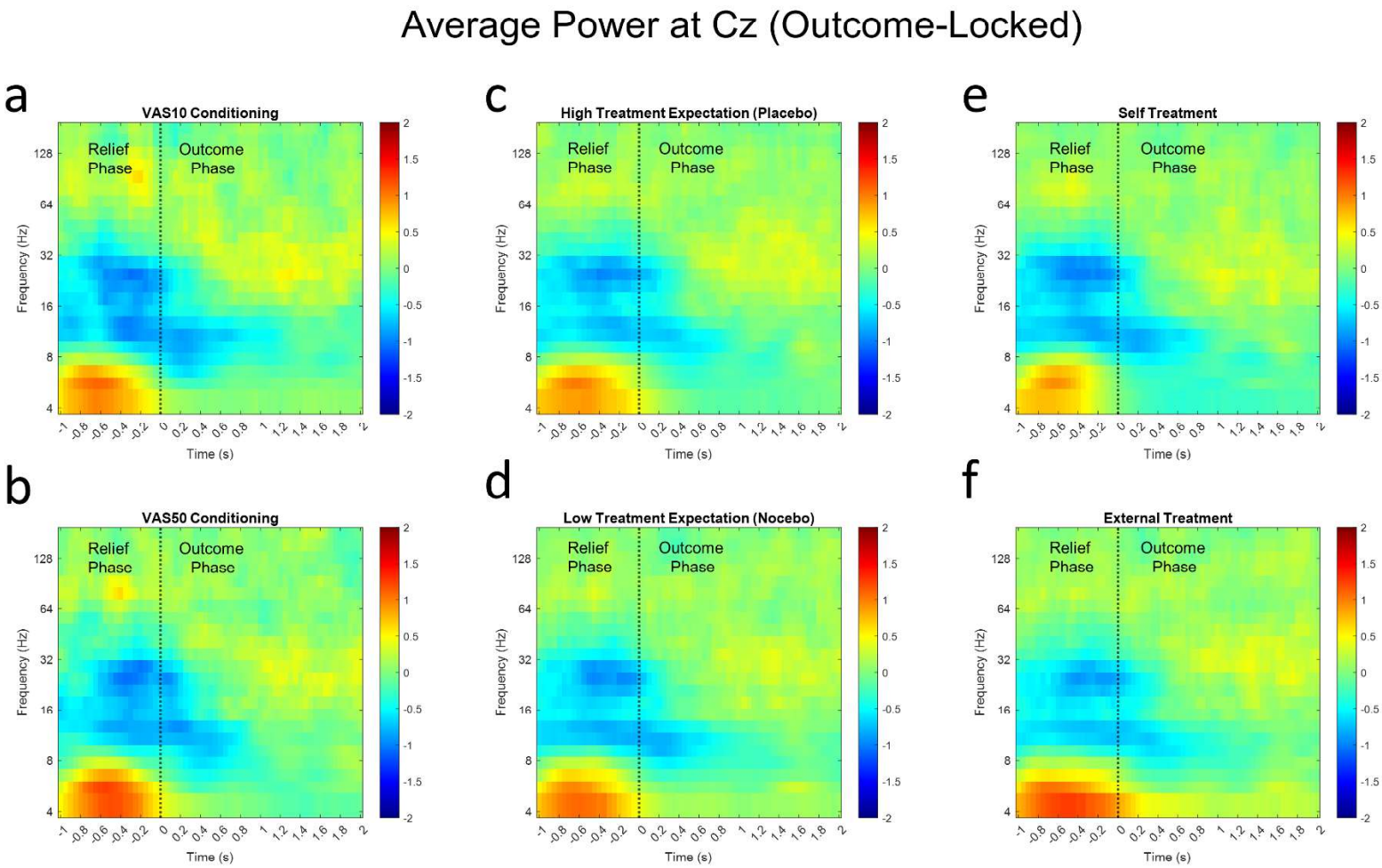

1195

1196 **Supplementary Figure 4.** Time-frequency plots represent averaged, baseline-corrected, cue-locked power
1197 values for each condition at Pz. (a) VAS10 versus (b) VAS50 conditioning, (c) high versus (d) low treatment
1198 expectation and (e) self- versus (f) external treatment. Warm colors represent positive *z*-values (increase
1199 of power compared to baseline) and cold colors represent negative *z*-values (decrease of power compared
1200 to baseline).

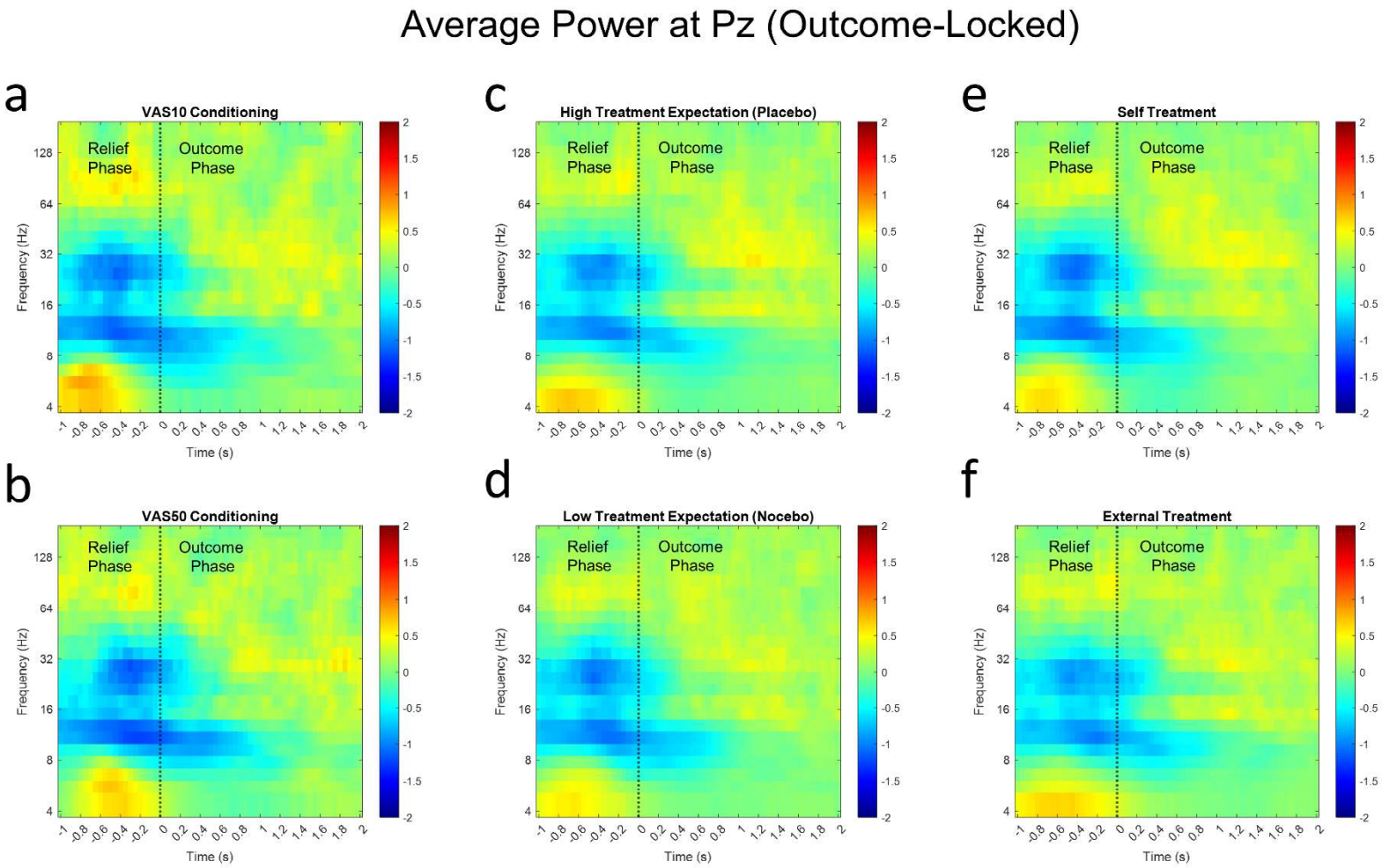
1201



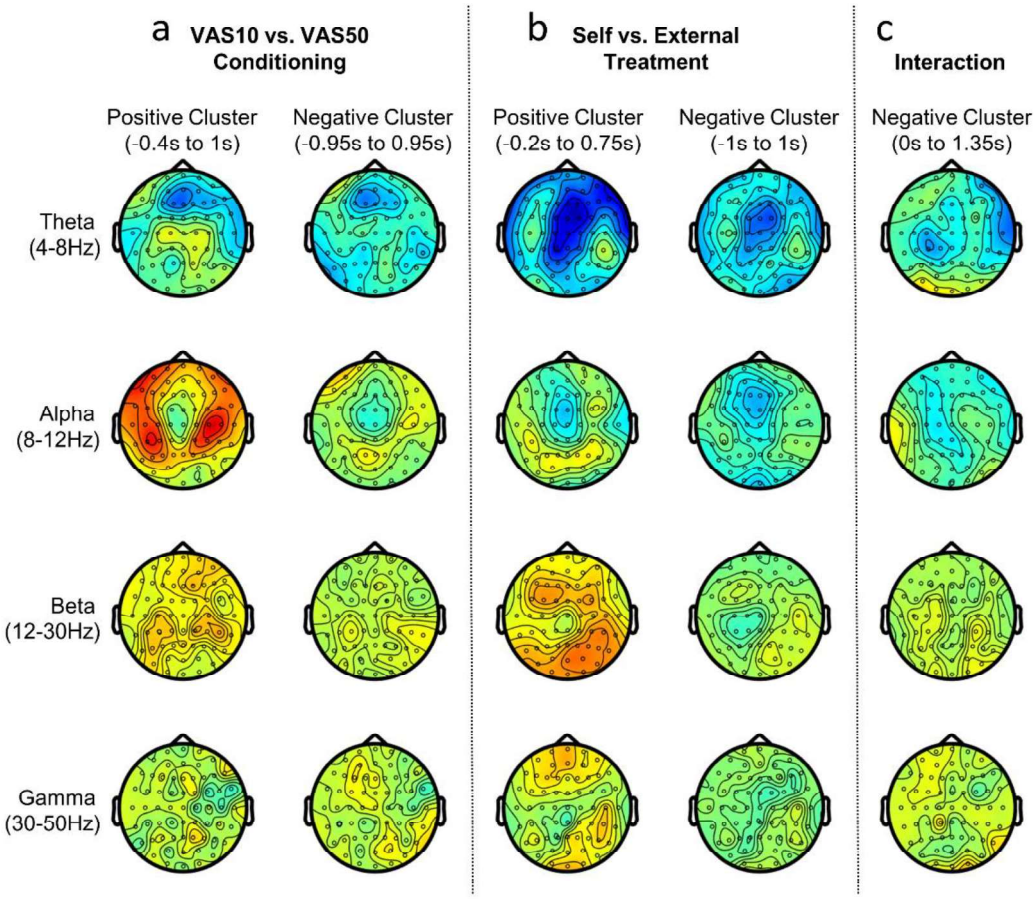

**Supplementary Figure 5.** Time-frequency plots represent averaged, baseline-corrected, outcome-locked power values for each condition at Fz. (a) VAS10 versus (b) VAS50 conditioning, (c) high versus (d) low treatment expectation and (e) self- versus (f) external treatment. Warm colors represent positive *z*-values (increase of power compared to baseline) and cold colors represent negative *z*-values (decrease of power compared to baseline).




**Supplementary Figure 6.** Time-frequency plots represent averaged, baseline-corrected, outcome-locked power values for each condition at Cz. (a) VAS10 versus (b) VAS50 conditioning, (c) high versus (d) low treatment expectation and (e) self- versus (f) external treatment. Warm colors represent positive *z*-values (increase of power compared to baseline) and cold colors represent negative *z*-values (decrease of power compared to baseline).

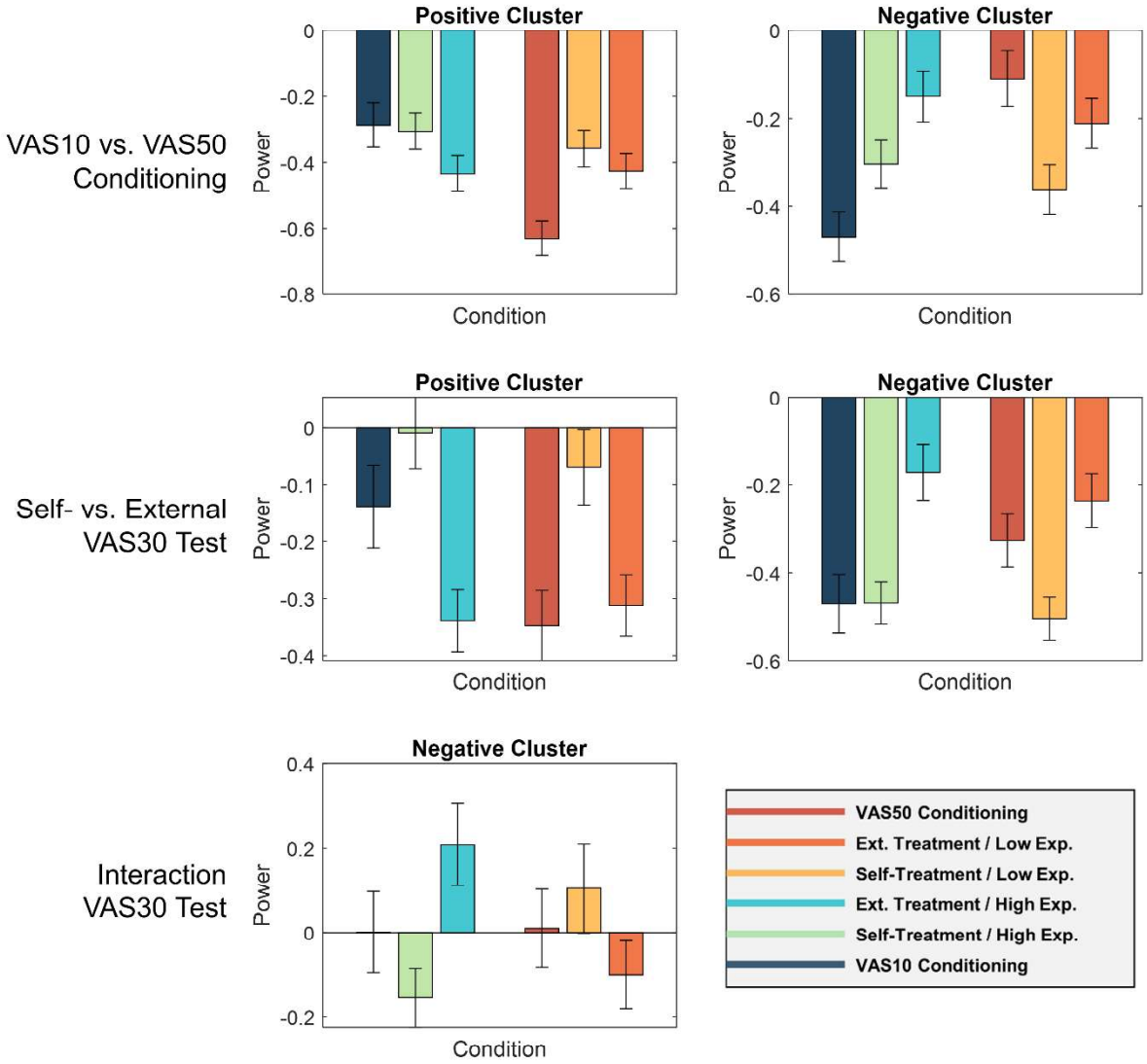



**Supplementary Figure 7.** Time-frequency plots represent averaged, baseline-corrected, outcome-locked power values for each condition at Pz. (a) VAS10 versus (b) VAS50 conditioning, (c) high versus (d) low treatment expectation and (e) self- versus (f) external treatment. Warm colors represent positive *z*-values (increase of power compared to baseline) and cold colors represent negative *z*-values (decrease of power compared to baseline).

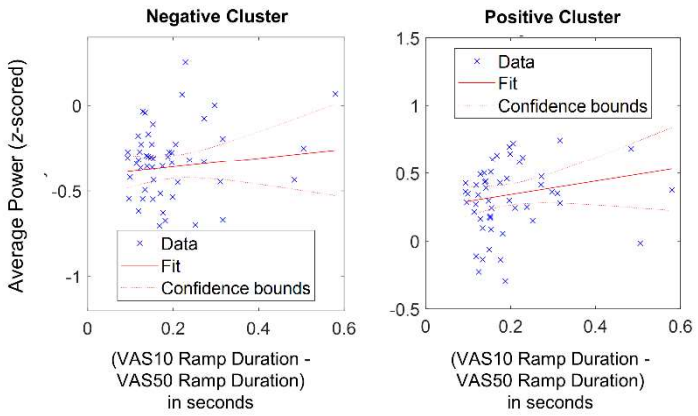



1223

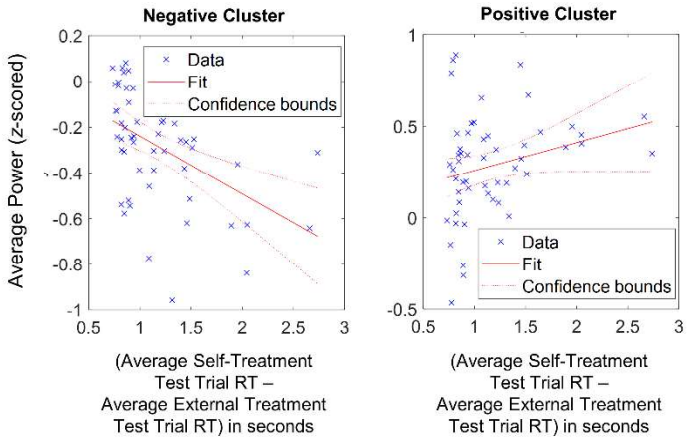
1224 **Supplementary Figure 8.** Topographies represent averaged *t*-values of pre-defined frequency bands
1225 (Theta 4-8Hz, Alpha 8-12Hz, Beta 12-30Hz and Low Gamma 30-50Hz) over the time range of significant
1226 clusters of (a) VAS10 versus VAS50 conditioning, (b) self- versus external treatment and (c) the interaction
1227 of agency and expectation.



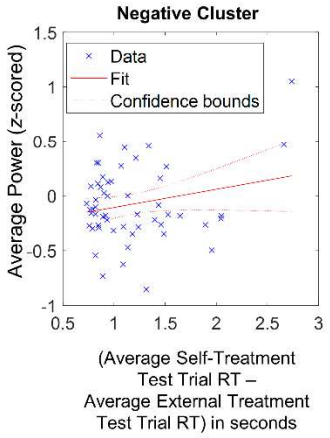

**Supplementary Figure 9.** Bar graphs represent averaged, baseline-corrected power values for each condition averaged over all data points included in significant clusters of (a) VAS10 versus VAS50 conditioning, (b) self- versus external treatment and (c) the interaction of agency and expectation. Error bars represent *SEM*.




**Supplementary Figure 10.** Linear regression of VAS10 versus VAS50 EEG clusters and individual ramp period differences, (left) of the negative cluster (outcome-locked; 0.95 to 0.95s; 4-48Hz; $R^2$ = 0.011, $F$(1,52) = 0.6, $p$ = 0.442), and (right) of the positive cluster (outcome-locked; -0.4 to 1s; 8-64Hz; $R^2$ = 0.0316, $F$(1,52) = 1.66, $p$ = 0.203). Here, each significant EEG cluster associated with differences in outcome temperatures (i.e. VAS10 versus VAS50 conditioning) was tested in a linear regression with the individual ramp time as a predictor. Each cross represents a single participant, the full red line represents the fitted regression line and dashed red lines represent 95% confidence bounds.




**Supplementary Figure 11.** Linear regression of self- versus external test trials EEG clusters and reaction time differences, (left) of the negative cluster (outcome-locked; -1 to 1s; 4-56Hz) and (right) of the positive cluster (outcome-locked; -0.2 to 0.75s; 11-54Hz). Here, each significant EEG cluster associated with differences in agency was tested in a linear regression with the individual reaction time difference between self- and external trials as a predictor. Each cross represents a single participant, the full red line represents the fitted regression line and dashed red lines represent 95% confidence bounds.




**Supplementary Figure 12.** Linear regression of the interaction EEG cluster and reaction time differences. Differences in reaction times are associated with longer outcome latencies in self-treatment. Here, the significant negative interaction EEG cluster (cue-locked; 0 to 1.35s; 4-13.5Hz) was tested in a linear regression with the individual reaction time difference between self- and external trials as a predictor. Each cross represents a single participant, the full red line represents the fitted regression line and dashed red lines represent 95% confidence bounds.

| | | | Agency Benefit | | Placebo Benefit | |
|---|---|---|---|---|---|---|
| Scale | *M* | *SD* | *Pearson's r* | *p-value* | *Pearson's r* | *p-value* |
| BDI-V | 20.74 | 15.31 | 0.25 | .06 | 0.18 | .18 |
| STAI-X2 (Trait) | 37.74 | 11.21 | 0.10 | .47 | 0.23 | .09 |
| STAI-X1 (State) | 38.77 | 11.47 | 0.16 | .26 | 0.15 | .15 |
| FKK-C | 13.94 | 6.67 | -0.04 | .77 | 0.08 | .55 |
| FKK-I | 24.41 | 5.94 | 0.15 | .27 | -0.07 | .62 |
| FKK-P | 15.39 | 6.36 | -0.02 | .91 | 0.23 | .09 |
| FKK-SK | 25.70 | 6.50 | -0.02 | .90 | -0.05 | .73 |
| LSHS_R | 7.20 | 6.60 | 0.02 | .91 | 0.22 | .12 |
| SWE | 20.69 | 4.28 | -0.08 | .61 | 0.01 | .95 |
| PCS | 17.56 | 9.21 | 0.15 | .28 | 0.10 | .46 |

**Supplementary Table 1.** Correlation of questionnaire data with individual agency and placebo benefit scores.

| Name | Estimate | SE | *t*-value | DF | *p*-Value | Lower Bound | Upper Bound |
|---|---|---|---|---|---|---|---|
| Intercept | 38.29 | 2.337 | 16.394 | 2588 | <.001 | 33.71 | 42.87 |
| Positive Agency Cluster (-0.2 to 0.75s; 11-54Hz) | -0.434 | 0.378 | -1.149 | 2588 | .261 | -1.175 | 0.307 |
| Negative Interaction Cluster (0 to 1.35s; 4-13.5Hz) | 0.739 | 0.294 | 2.510 | 2588 | .0121* | 0.162 | 1.316 |

**Supplementary Table 2.** Fixed effects coefficients of the trial-by-trial LME model with 95% confidence intervals (lower bound and upper bound) for each cluster and intercept for VAS outcome ratings.

## References


1. Tracey, I., and Mantyh, P.W. (2007). The cerebral signature for pain perception and its modulation. Neuron *55*, 377–391. 10.1016/j.neuron.2007.07.012.

2. Beck, B., Di Costa, S., and Haggard, P. (2017). Having control over the external world increases the implicit sense of agency. Cognition *162*, 54–60. 10.1016/j.cognition.2017.02.002.

3. Helmchen, C., Mohr, C., Erdmann, C., Binkofski, F., and Büchel, C. (2006). Neural activity related to self- versus externally generated painful stimuli reveals distinct differences in the lateral pain system in a parametric fMRI study. Hum. Brain Mapp. *27*, 755–765. 10.1002/hbm.20217.

4. Jensen, M.P., and Karoly, P. (1991). Control beliefs, coping efforts, and adjustment to chronic pain. J. Consult. Clin. Psychol. *59*, 431–438. 10.1037//0022-006x.59.3.431.

5. Kakigi, R., and Shibasaki, H. (1992). Mechanisms of pain relief by vibration and movement. J. Neurol. Neurosurg. Psychiatry *55*, 282–286. 10.1136/jnnp.55.4.282.

6. Karsh, N., Goldstein, O., and Eitam, B. (2018). Evidence for pain attenuation by the motor system-based judgment of agency. Conscious. Cogn. *57*, 134–146. 10.1016/j.concog.2017.11.012.

7. Mohr, C., Leyendecker, S., and Helmchen, C. (2008). Dissociable neural activity to self- vs. externally administered thermal hyperalgesia: a parametric fMRI study. Eur. J. Neurosci. *27*, 739–749. 10.1111/j.1460-9568.2008.06036.x.

8. Mohr, C., Leyendecker, S., Petersen, D., and Helmchen, C. (2012). Effects of perceived and exerted pain control on neural activity during pain relief in experimental heat hyperalgesia: a fMRI study. Eur. J. Pain Lond. Engl. *16*, 496–508. 10.1016/j.ejpain.2011.07.010.

9. Müller, M., and Netter, P. (2000). Relationship of subjective helplessness and pain perception after electric skin stimuli. Stress Med. *16*, 109–115. 10.1002/(SICI)1099-1700(200003)16:2<109::AID-SMI837>3.0.CO;2-#.

10. Pellino, T.A., and Ward, S.E. (1998). Perceived control mediates the relationship between pain severity and patient satisfaction. J. Pain Symptom Manage. *15*, 110–116.

11. Pervin, L.A. (1963). The need to predict and control under conditions of threat. J. Pers. *31*, 570–587. 10.1111/j.1467-6494.1963.tb01320.x.

12. Staub, E., Tursky, B., and Schwartz, G.E. (1971). Self-control and predictability: their effects on reactions to aversive stimulation. J. Pers. Soc. Psychol. *18*, 157–162. 10.1037/h0030851.

13. Thompson, S.C. (1981). Will it hurt less if I can control it? A complex answer to a simple question. Psychol. Bull. *90*, 89–101. 10.1037/0033-2909.90.1.89.

14. Wang, Y., Wang, J.-Y., and Luo, F. (2011). Why Self-Induced Pain Feels Less Painful than Externally Generated Pain: Distinct Brain Activation Patterns in Self- and Externally Generated Pain. PLOS ONE *6*, e23536. 10.1371/journal.pone.0023536.

15. Weisenberg, M., Wolf, Y., Mittwoch, T., Mikulincer, M., and Aviram, O. (1985). Subject versus experimenter control in the reaction to pain. Pain *23*, 187–200. 10.1016/0304-3959(85)90059-4.

16. Wiech, K., Kalisch, R., Weiskopf, N., Pleger, B., Stephan, K.E., and Dolan, R.J. (2006). Anterolateral Prefrontal Cortex Mediates the Analgesic Effect of Expected and Perceived Control over Pain. J. Neurosci. *26*, 11501–11509. 10.1523/JNEUROSCI.2568-06.2006.

17. Ballantyne, J.C., Carr, D.B., Chalmers, T.C., Dear, K.B., Angelillo, I.F., and Mosteller, F. (1993). Postoperative patient-controlled analgesia: meta-analyses of initial randomized control trials. J. Clin. Anesth. *5*, 182–193. 10.1016/0952-8180(93)90013-5.

18. McNicol, E.D., Ferguson, M.C., and Hudcova, J. (2015). Patient controlled opioid analgesia versus non-patient controlled opioid analgesia for postoperative pain. Cochrane Database Syst. Rev. 10.1002/14651858.CD003348.pub3.

19. Blakemore, S.-J., Wolpert, D.M., and Frith, C.D. (1998). Central cancellation of self-produced tickle sensation. Nat. Neurosci. *1*, 635–640. 10.1038/2870.

20. Blakemore, S.-J., Frith, C.D., and Wolpert, D.M. (1999). Spatio-Temporal Prediction Modulates the Perception of Self-Produced Stimuli. J. Cogn. Neurosci. *11*, 551–559. 10.1162/089892999563607.

21. Claxton, G. (1975). Why can't we tickle ourselves? Percept. Mot. Skills *41*, 335–338. 10.2466/pms.1975.41.1.335.

22. Weiskrantz, L., Elliott, J., and Darlington, C. (1971). Preliminary Observations on Tickling Oneself. Nature *230*, 598–599. 10.1038/230598a0.

23. Anchisi, D., and Zanon, M. (2015). A Bayesian Perspective on Sensory and Cognitive Integration in Pain Perception and Placebo Analgesia. PLOS ONE *10*, e0117270. 10.1371/journal.pone.0117270.

24. Büchel, C., Geuter, S., Sprenger, C., and Eippert, F. (2014). Placebo analgesia: a predictive coding perspective. Neuron *81*, 1223–1239. 10.1016/j.neuron.2014.02.042.

25. Ongaro, G., and Kaptchuk, T.J. (2019). Symptom perception, placebo effects, and the Bayesian brain. Pain *160*, 1–4. 10.1097/j.pain.0000000000001367.

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

27. Grahl, A., Onat, S., and Büchel, C. (2018). The periaqueductal gray and Bayesian integration in placebo analgesia. eLife *7*, e32930. 10.7554/eLife.32930.

28. Ernst, M.O., and Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. Nature *415*, 429–433. 10.1038/415429a.

29. Darwin, C. (1872). The Expression of Emotions in Man and Animals (John Murray).

30. Blakemore, S.-J., Wolpert, D., and Frith, C. (2000). Why can't you tickle yourself? NeuroReport *11*, R11.

31. Brown, H., Adams, R.A., Parees, I., Edwards, M., and Friston, K. (2013). Active inference, sensory attenuation and illusions. Cogn. Process. *14*, 411–427. 10.1007/s10339-013-0571-3.

32. Nickel, M.M., Tiemann, L., Hohn, V.D., May, E.S., Gil Ávila, C., Eippert, F., and Ploner, M. (2022). Temporal–spectral signaling of sensory information and expectations in the cerebral processing of pain. Proc. Natl. Acad. Sci. *119*, e2116616119. 10.1073/pnas.2116616119.

33. Tiemann, L., May, E.S., Postorino, M., Schulz, E., Nickel, M.M., Bingel, U., and Ploner, M. (2015). Differential neurophysiological correlates of bottom-up and top-down modulations of pain. PAIN *156*, 289–296. 10.1097/01.j.pain.0000460309.94442.44.

34. Strube, A., Rose, M., Fazeli, S., and Büchel, C. (2021). The temporal and spectral characteristics of expectations and prediction errors in pain and thermoception. eLife *10*, e62809. 10.7554/eLife.62809.

35. Rigoux, L., Stephan, K.E., Friston, K.J., and Daunizeau, J. (2014). Bayesian model selection for group studies - revisited. NeuroImage *84*, 971–985. 10.1016/j.neuroimage.2013.08.065.

36. Stephan, K.E., Penny, W.D., Daunizeau, J., Moran, R.J., and Friston, K.J. (2009). Bayesian model selection for group studies. NeuroImage *46*, 1004–1017. 10.1016/j.neuroimage.2009.03.025.

37. Fazeli, S., and Büchel, C. (2018). Pain-Related Expectation and Prediction Error Signals in the Anterior Insula Are Not Related to Aversiveness. J. Neurosci. *38*, 6461. 10.1523/JNEUROSCI.0671-18.2018.

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

39. Geuter, S., and Buchel, C. (2013). Facilitation of Pain in the Human Spinal Cord by Nocebo Treatment. J. Neurosci. *33*, 13784–13790. 10.1523/JNEUROSCI.2191-13.2013.

40. Hird, E.J., Jones, A.K.P., Talmi, D., and El-Deredy, W. (2018). A comparison between the neural correlates of laser and electric pain stimulation and their modulation by expectation. J. Neurosci. Methods *293*, 117–127. 10.1016/j.jneumeth.2017.09.011.

41. Woo, C.-W., Schmidt, L., Krishnan, A., Jepma, M., Roy, M., Lindquist, M.A., Atlas, L.Y., and Wager, T.D. (2017). Quantifying cerebral contributions to pain beyond nociception. Nat. Commun. *8*, 14211. 10.1038/ncomms14211.

42. Ploner, M., Sorg, C., and Gross, J. (2017). Brain Rhythms of Pain. Trends Cogn. Sci. *21*, 100–110. 10.1016/j.tics.2016.12.001.

43. Hu, L., Peng, W., Valentini, E., Zhang, Z., and Hu, Y. (2013). Functional features of nociceptive-induced suppression of alpha band electroencephalographic oscillations. J. Pain *14*, 89–99. 10.1016/j.jpain.2012.10.008.

44. May, E.S., Butz, M., Kahlbrock, N., Hoogenboom, N., Brenner, M., and Schnitzler, A. (2012). Pre- and post-stimulus alpha activity shows differential modulation with spatial attention during the processing of pain. NeuroImage *62*, 1965–1974. 10.1016/j.neuroimage.2012.05.071.

45. Mouraux, A., Guérit, J.M., and Plaghki, L. (2003). Non-phase locked electroencephalogram (EEG) responses to CO2 laser skin stimulations may reflect central interactions between A partial partial differential- and C-fibre afferent volleys. Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol. *114*, 710–722. 10.1016/s1388-2457(03)00027-0.

46. Ploner, M., Gross, J., Timmermann, L., Pollok, B., and Schnitzler, A. (2006). Pain suppresses spontaneous brain rhythms. Cereb. Cortex N. Y. N 1991 *16*, 537–540. 10.1093/cercor/bhj001.

47. Zhang, Z.G., Hu, L., Hung, Y.S., Mouraux, A., and Iannetti, G.D. (2012). Gamma-band oscillations in the primary somatosensory cortex--a direct and obligatory correlate of subjective pain intensity. J. Neurosci. Off. J. Soc. Neurosci. *32*, 7429–7438. 10.1523/JNEUROSCI.5877-11.2012.

48. Bingel, U., Schoell, E., Herken, W., Büchel, C., and May, A. (2007). Habituation to painful stimulation involves the antinociceptive system. Pain *131*, 21–30. 10.1016/j.pain.2006.12.005.

49. Greffrath, W., Baumgärtner, U., and Treede, R.-D. (2007). Peripheral and central components of habituation of heat pain perception and evoked potentials in humans. PAIN *132*, 301–311. 10.1016/j.pain.2007.04.026.

50. Jepma, M., Jones, M., and Wager, T.D. (2014). The Dynamics of Pain: Evidence for Simultaneous Site-Specific Habituation and Site-Nonspecific Sensitization in Thermal Pain. J. Pain *15*, 734–746. 10.1016/j.jpain.2014.02.010.

51. Jepma, M., and Wager, T.D. (2013). Multiple potential mechanisms for context effects on pain. Pain *154*, 629–631. 10.1016/j.pain.2013.02.011.

52. Rennefeld, C., Wiech, K., Schoell, E.D., Lorenz, J., and Bingel, U. (2010). Habituation to pain: Further support for a central component. PAIN® *148*, 503–508. 10.1016/j.pain.2009.12.014.

53. Colloca, L., Tinazzi, M., Recchia, S., Le Pera, D., Fiaschi, A., Benedetti, F., and Valeriani, M. (2008). Learning potentiates neurophysiological and behavioral placebo analgesic responses. Pain *139*, 306–314. 10.1016/j.pain.2008.04.021.

54. Morton, D.L., El-Deredy, W., Watson, A., and Jones, A.K.P. (2010). Placebo analgesia as a case of a cognitive style driven by prior expectation. Brain Res. *1359*, 137–141. 10.1016/j.brainres.2010.08.046.

55. Wager, T.D., Matre, D., and Casey, K.L. (2006). Placebo effects in laser-evoked pain potentials. Brain. Behav. Immun. *20*, 219–230. 10.1016/j.bbi.2006.01.007.

56. Jepma, M., Koban, L., van Doorn, J., Jones, M., and Wager, T.D. (2018). Behavioural and neural evidence for self-reinforcing expectancy effects on pain. Nat. Hum. Behav. *2*, 838–855. 10.1038/s41562-018-0455-8.

57. Eckert, A.-L., Pabst, K., and Endres, D.M. (2022). A Bayesian model for chronic pain. Front. Pain Res. *3*.

58. Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behav. Res. Methods *41*, 1149–1160. 10.3758/BRM.41.4.1149.

59. Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences 2nd ed. (Routledge) 10.4324/9780203771587.

60. Schenk, L.A., Sprenger, C., Onat, S., Colloca, L., and Büchel, C. (2017). Suppression of Striatal Prediction Errors by the Prefrontal Cortex in Placebo Hypoalgesia. J. Neurosci. *37*, 9715–9723. 10.1523/JNEUROSCI.1101-17.2017.

61. Thomaidou, M.A., Veldhuijzen, D.S., Meulders, A., and Evers, A.W.M. (2021). An experimental investigation into the mediating role of pain-related fear in boosting nocebo hyperalgesia. Pain *162*, 287–299. 10.1097/j.pain.0000000000002017.

62. van de Sand, M.F., Menz, M.M., Sprenger, C., and Büchel, C. (2018). Nocebo-induced modulation of cerebral itch processing – An fMRI study. NeuroImage *166*, 209–218. 10.1016/j.neuroimage.2017.10.056.

63. Bentall, R.P., and Slade, P.D. (1985). Reliability of a scale measuring disposition towards hallucination: a brief report. Personal. Individ. Differ. *6*, 527–529. 10.1016/0191-8869(85)90151-5.

64. Lincoln, T.M., Keller, E., and Rief, W. (2009). Die Erfassung von Wahn und Halluzinationen in der Normalbevölkerung. Diagnostica *55*, 29–40. 10.1026/0012-1924.55.1.29.

65. Lincoln, T.M., and Keller, E. (2008). Delusions and hallucinations in students compared to the general population. Psychol. Psychother. Theory Res. Pract. *81*, 231–235. 10.1348/147608308X297096.

66. Laux, L. (1981). Das State-Trait-Angstinventar (STAI) : theoretische Grundlagen und Handanweisung (Beltz).

67. Krampen, G. (1991). Fragebogen zu Kompetenz- und Kontrollüberzeugungen: (FKK) (Hogrefe, Verl. für Psychologie).

68. Jerusalem, M., and Schwarzer, R. (2003). SWE - Skala zur Allgemeinen Selbstwirksamkeitserwartung. 10.23668/PSYCHARCHIVES.4515.

69. Schwarzer, R., and Jerusalem, M. (1999). Skala zur Erfassung von Lehrer-und Schülermerkmalen. Dok. Psychom. Verfahr. Im Rahm. Wiss. Begleit. Modellvers. Selbstwirksame Schulen.

70. Meyer, K., Sprott, H., and Mannion, A.F. (2008). Cross-cultural adaptation, reliability, and validity of the German version of the Pain Catastrophizing Scale. J. Psychosom. Res. *64*, 469–478. 10.1016/j.jpsychores.2007.12.004.

71. Sullivan, M., Bishop, S., and Pivik, J. (1995). The Pain Catastrophizing Scale: Development and validation. Psychol. Assess. *7*, 524–532. 10.1037/1040-3590.7.4.524.

72. Klem, G.H., Lüders, H.O., Jasper, H.H., and Elger, C. (1999). The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. Electroencephalogr. Clin. Neurophysiol. Suppl. *52*, 3–6.

73. Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. *2011*, 156869. 10.1155/2011/156869.

74. Hipp, J.F., Engel, A.K., and Siegel, M. (2011). Oscillatory synchronization in large-scale cortical networks predicts perception. Neuron *69*, 387–396. 10.1016/j.neuron.2010.12.027.

75. Jung, T.P., Makeig, S., Humphries, C., Lee, T.W., McKeown, M.J., Iragui, V., and Sejnowski, T.J. (2000). Removing electroencephalographic artifacts by blind source separation. Psychophysiology *37*, 163–178.

76. Makeig, S., Bell, A., Jung, T.-P., and Sejnowski, T. (1996). Independent Component Analysis of Electroencephalographic Data. *8*.

77. Grandchamp, R., and Delorme, A. (2011). Single-trial normalization for event-related spectral decomposition reduces sensitivity to noisy trials. Front. Psychol. *2*, 236. 10.3389/fpsyg.2011.00236.

78. Daunizeau, J., Adam, V., and Rigoux, L. (2014). VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data. PLOS Comput. Biol. *10*, e1003441. 10.1371/journal.pcbi.1003441.

79. Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods *164*, 177–190. 10.1016/j.jneumeth.2007.03.024.

## 16.0 Eidesstattliche Versicherung

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe.

Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.

Ich erkläre mich einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Unterschrift: .....