Explaining the Placebo Effect: Aliefs, Beliefs, and Conditioning

[Word Count: 9757, including notes and references]
Forthcoming in *Philosophical Psychology*
Matthew C. Haug
The College of William & Mary
mchaug@wm.edu

Abstract

There are a number of competing psychological accounts of the placebo effect, and much of the recent debate centers on the relative importance of classical conditioning and conscious beliefs. In this paper, I discuss apparent problems with these accounts and with “disjunctive” accounts that deny that placebo effects can be given a unified psychological explanation. The fact that some placebo effects seem to be mediated by cognitive states with content that is consciously inaccessible and inferentially isolated from a subject’s beliefs motivates an account of the placebo effect in terms of subdoxastic cognitive states. I propose that *aliefs*, subdoxastic cognitive states that are associative, automatic, and arational, can provide a unified psychological account of the placebo effect. This account also has the potential to illuminate interesting connections to other psychological phenomena.

0. Introduction

The placebo effect has become a hot topic in psychology, psychiatry, neuroscience, and brain imaging. (See Benedetti (2008) and Price et al. (2008) for recent reviews.) Much of the best clinical and experimental work has focused on placebo analgesia, although good evidence exists that placebo effects occur for a wide variety of conditions: from Parkinson’s disease to immune responses.

Recent research suggests that there is an interesting and important range of placebo effects that cannot be explained by appealing to phenomena that have historically been lumped
together with placebos, such as regression to the mean, the natural history (untreated progression) of the condition, or response biases. Rather, the psychosocial context in which treatment takes place has an important effect on treatment outcomes. For many conditions, mere simulation of a therapy can produce effects similar to those produced by the therapy itself.¹

There are several competing theoretical accounts of the placebo effect, with much recent emphasis placed on the relative roles of classical conditioning and conscious beliefs (Hoffman et al., 2005, p. 257; Stewart-Williams & Podd, 2004). After discussing these accounts in Section 1, I discuss some cases of the placebo effect that are not handled well by each of them (Section 2). One way to respond to such problem cases is to deny that placebo effects can be given a unified psychological explanation. In Section 3, I point out some general theoretical problems with such “disjunctive” strategies, and I identify specific problems with any disjunctive account that includes conditioning and belief-based accounts as distinct components. I then use some of these problems to motivate a new account in Section 4, which claims that placebo effects are in general mediated by subdoxastic cognitive states that Tamar Gendler (2008(a), 2008(b)) has called “aliefs”.² Not only does this account provide a better explanation of the problem cases than its competitors, it also allows us to improve and refine our theory of alief and illuminate interesting connections to other psychological phenomena (Sections 5 and 6).

1. Prominent Accounts of the Placebo Effect
1.1. Classical Conditioning

Since the 1960s, many experiments have shown that a wide variety of physiological and behavioral responses in both humans and non-human animals are subject to conditioning (e.g. Herrnstein, 1962; Ader & Cohen, 1982; Goebel et al., 2002). That is, subjects exhibit the same kind of responses (e.g. suppression of the immune system or of learned behavior) to conditioned
stimuli (e.g. saline injections) as they do to unconditioned stimuli (e.g. cyclophosphamide or scopolamine injections).

This has led some authors to propose that the placebo response, in general, can be explained as an instance of classical conditioning. For example, according to Wickramasekera (1980) placebo responses are brought about through the association of neutral objects (e.g., the hospital, the doctor, pills, injections) with effective treatment. After repeated association, these conditioned stimuli may themselves produce a therapeutic effect. Administration of the active drug (either in an experimental setting or in the normal course of an individual’s life) is the unconditioned stimulus, which produces the unconditioned response of improvement in the given condition. Relevant aspects of the environmental context (aside from the active drug) are the conditioned stimuli, which produce the conditioned response of improvement of the given illness or condition (which is usually of lesser magnitude than the unconditioned response).

1.2. Expectancy, Belief, and Desire

After the cognitive revolution in psychology, behaviorist paradigms like Pavlovian, classical conditioning fell out of favor. Concomitantly, a variety of cognitive theories of the placebo effect rose to prominence. These cognitive theories hold that placebo effects result from beliefs about the efficacy of the treatment, expectations about one’s reaction to treatment (perceived likelihood of a given outcome), or desires that one’s condition improve. Particular cognitive theories emphasize one or more of these cognitive states, perhaps together with other relatively high-level psychological states like emotions and memories. For instance, Irving Kirsch (1997, 2010) focuses on “response expectancies” – the subject’s consciously accessible beliefs about her own automatic, involuntary reactions to particular situational cues – and argues that they are the most important causes of placebo effects and are required for most placebo
effects in humans. Price & Fields (1997) argue that a desire for pain relief mediates placebo analgesia in addition to expectations about the efficacy of a particular treatment. They claim that the addition of desire as a variable accounts for the increase in placebo response with the severity of clinical pain and for the fact that placebo responses in experimental situations increase as a function of the pain’s duration and severity (ibid., p. 128). To take one more example, several groups of researchers have proposed that the placebo effect is a special case of reward processing, which involves motivational and affective-emotional states as well as expectations. For example, de le Fuente-Fernandez et al. (2001) argue that the placebo effect for Parkinson’s disease is mediated by expectations of reward (specifically, the expectation of therapeutic benefit). In addition, Petrovic et al. (2005) provide brain-imaging evidence that suggests that “placebo analgesia is a special case of reward processing and that placebo treatment could modulate emotional perception in the same way as [it modulates] pain perception” (ibid., p. 957).

2. Problems Facing These Accounts

It is well-known that none of these accounts is general enough to cover all placebo effects. In Sections 2.1 and 2.2, I discuss cases that pose apparent problems for conditioning and belief-based accounts, respectively. In Section 2.3, I discuss placebo effects related to the “formal” features of a treatment, which do not seem to be well-handled by either account.

2.1. Problem Cases for Conditioning Accounts

A number of experimental studies pose problems for conditioning accounts. Montgomery & Kirsch’s (1997) study suggests that consciously accessible expectations are more important than conditioning. A painful electrical stimulus (calibrated to adjust for individual differences) (level 6 out of 10) was administered to participants’ forearms. Some of these shocks were given with a placebo ointment (a mixture of iodine, oil of thyme, and water, which participants were
told was a new topical analgesic called “Trivaricane”), while others were not. This established a “baseline,” pre-manipulation placebo effect for the subjects (a mean pain reduction of 0.29). Then, before a series of manipulation trials, one group of subjects was told that “the intensity would be reduced on medication trials to examine the effectiveness of Trivaricane at lower intensities” (ibid., p. 110); that is, they were informed that the stimulus intensity would be reduced only when the ointment was applied. The other, “uninformed” group was not given this information; that is, for this group, the intensity of the stimulus was surreptitiously lowered (to level 3, the same in both groups) when paired with the placebo ointment. For both groups, the stimulus remained unchanged (level 6) when no ointment was applied. Following the manipulation trials, all subjects were asked what they expected the pain intensity for shocks (at the original stimulus level) to be with and without Trivaricane in a subsequent series of trials. In the informed group, the mean expected pain intensity for shocks with Trivaricane was higher (4.92) than in the uninformed group (2.75). Mean expected pain intensity for shocks without Trivaricane was approximately equal in the two groups (6.83 and 6.92). In the posttest trials, each group was administered a series of shocks at the original level 6; some shocks were paired with the placebo ointment and some were not. Members of the group that had been informed of the intensity reduction had a lower analgesic response to the placebo ointment than members of the uninformed group. The conditioning model of placebo analgesia predicts that changes in conscious expectation should have little effect on pain reports, contrary to these results.³

In another experiment, Benedetti et al. (2003) gave some study participants a saline injection and told them the shot would intensify their pain; other volunteers were also given the placebo pain promoter but in addition underwent conditioning to decrease pain in which the saline shot was preceded by injections of the non-steroidal anti-inflammatory drug (NSAID)
ketorolac. In both groups pain increased, demonstrating that negative expectation is a powerful nocebo (“negative placebo”) in the case of pain. Further, anticipating more pain led to increased levels of pain despite conditioning to an analgesic, apparently showing that expectation influences pain more than conditioning does.

2.2. Problems Facing Belief-based Accounts

However, recent studies have also raised trouble for extant cognitive theories. For example, Benedetti et al. (1998) investigated the effects of prior opioid exposure on placebo analgesia and placebo respiratory depression. Participants in the study received buprenorphine to control postoperative pain, which also decreases respiratory rate and volume (although in low doses not at levels that are readily detectable by the subject). Initial pain intensity and respiratory pattern (breath volume and rate) was measured and subjects were then given buprenorphine until satisfactory analgesia (pain level ≤ 4) occurred. Twelve to eighteen hours after surgery, when pain levels had increased, respiratory pattern was again assessed and patients were given either buprenorphine or a saline solution (which they were told was buprenorphine) in a double-blind design. (A no treatment group was included for comparison.)

The researchers concluded that “a placebo respiratory depressant response is more pronounced if an opiate respiratory depressant response was present the day before. In addition, placebo respiratory depression is completely independent of both buprenorphine analgesia and placebo analgesia, indicating that the amount of pain reduction is not important for the occurrence of the placebo respiratory response” (ibid., p. 317). They continue:

In fact, the mild respiratory depressant effect of buprenorphine was completely unnoticed by our patients, thus suggesting that cognition and probably anxiety are not important. In addition, since the placebo respiratory depressant response appears to be completely independent of the analgesic effect (and thus of cognitive and affective influences), a conditioning mechanism not involving
cognition seems to be in accordance with the conditioning hypothesis. (ibid., p. 318)

In another portion of Benedetti et al.’s (2003) study, participants were told that a saline shot would alter levels (either up or down, depending on the group) of growth hormone or the stress hormone cortisol. (The study employed a double-blind design and included a no treatment group for comparison.) However, the suggestions had no effect on either hormone. By contrast, a saline injection did alter hormone concentrations when the researchers had previously conditioned subjects with sumatriptan, a drug used to treat migraines that influences secretion of growth hormone and cortisol. These placebo-induced biological changes occurred even if the participants were told the saline injection would have an effect opposite to that of sumatriptan. Thus, it seems that conditioning can manipulate involuntary physiological processes more than beliefs can.

2.3. “Formal” Features of Treatments and the Placebo Effect

A smattering of intriguing studies supports the claim that the form in which a treatment is administered has important therapeutic implications; the color of pills, number of doses, mechanism of delivery, and contextual properties of a treatment, such as price and brand prestige, may each play a role in the placebo effect. (Moerman (2002, Ch. 5) provides an interesting summary of many of these phenomena.) For example, De Craen and colleagues’ (1996) review of studies on the effects of the color of pills finds that people perceive red, yellow, and orange to be associated with a stimulant effect, while blue and green are related to a tranquilizing effect. Further, the color of a drug does seem to influence its effectiveness: in one study, patients who took blue pills fell asleep faster and slept longer than patients taking orange pills. (However, consistent trends were not apparent in all the studies examined.) De Craen et al. (1999) performed a meta-analysis of the placebo arms of ulcer medication and discovered that
patients who took a placebo four times a day had healing rates that were 6% to 8% higher than patients who took the same placebo two times a day. In another meta-analysis, De Craen et al. (2000) examined twenty-two placebo trials of sumatriptan, to investigate whether administering the drug subcutaneously by injection or orally by pill was more effective. As in the previous study, there was a small, but highly statistically significant, difference: subcutaneous placebos seem to be more effective than oral placebos.

In a recent study, Waber et al. (2008), investigated the effect of price on the effectiveness of a placebo. All participants were told that the placebo pills they were given were similar to codeine; half were told that each pill cost $2.50, and the other half were told that the price per pill had been discounted to $0.10 per pill. Using a standard electric shock protocol for studying pain, they found that approximately 20% more patients in the “regular-price” group experienced pain reduction after taking the pill than patients in the “low-price” group. “Considering all voltages tested, pain reduction was greater for the regular-price pill (P < .001). In addition, for 26 of 29 intensities (from 10 to 80 V), mean pain reduction was greater for the regular-price pill” (ibid., p. 1017). Finally, in a portion of Brarthwaite & Cooper’s (1981) study, one group of subjects was given placebo pills labeled “analgesic tablets.” Another group was given the same pills, but they were labeled with the brand name of one of the most popular aspirin-based analgesics in the UK. By one measure, the branded pills were 10% more effective in reducing headache pain than the unbranded placebos.

None of these results seems to favor either conditioning or cognitive accounts. One could account for these results in the conditioning paradigm by positing that subjects’ previous life experiences amounted to conditioning trials. For example, injections typically contain stronger doses of a drug than pills do. Since a stronger unconditioned stimulus leads to a stronger
conditioned response, one would expect the results obtained by de Craen et al. (1999).

Alternatively, from the cognitive perspective, one could argue that life experiences have led subjects to believe that injections will have stronger effects than pills, that more pills will be more effective than fewer pills, that warm colors are invigorating and cool colors are calming, and that more expensive items tend to be more effective than cheaper ones of the same kind.

While a proponent of either of these views may be happy with these explanations, they seem somewhat strained and artificial on closer examination. On the one hand, any “conditioning” that subjects received with respect to these factors throughout their normal lives would be highly irregular, intermittent and ambiguous. It is also hard to see how a simple conditioning account could explain the cross-cultural variation that de Craen and colleagues (2000) found in the injection studies; in the United States subcutaneous placebos were significantly more effective than oral placebos, but in Europe this difference was not as pronounced. One might try appealing to differences in the usage rates of pills and injections on the two continents, but it seems that abandoning the pure conditioning account and invoking cultural assumptions about the relative efficacy of pills and injections is more plausible. (For further discussion of cross-cultural variation see Moerman (2002, Ch. 6).)

On the other hand, it seems hard to believe that participants in these studies had explicit, conscious expectancies about the differences in efficacy related to differences in color, number, mechanism, or price. (For instance, did Branthwaite & Cooper’s subjects explicitly believe that branded pills were more effective than non-branded ones? It seems more likely that the increased confidence in obtaining relief with a well-known brand was represented by a different kind of psychological state.) At best, participants had implicit beliefs about these matters that they were not consciously aware of during the course of the studies. As discussed above, these results can
be forced into the conditioning or belief-based models. However, as I discuss below, I believe that the proposal offered in Section 4 offers a better account.

3. Disjunctive Accounts vs. the Virtues of Unification

One reply to these problems is to claim that the placebo effect is mediated by different kinds of psychological mechanisms in different cases. A simple way of doing so would be to offer a disjunctive, psychological account of the placebo effect. A more complex, and probably more realistic, alternative approach would be to invoke a complex, psychological mechanism with many variables – conditioned stimuli, expectancies, desires, emotional states – that could be modeled with structural equations or directed acyclic graphs. In some cases, such as those discussed above, specific variables would take on a value of zero and drop out of mechanism. Finally, one could take a more radical, eliminative line and claim that there is no such psychological phenomenon as the placebo effect. Rather, there is merely a cluster of unrelated biochemical mechanisms (cf. Benedetti, 2008; Price et al., 2008, p. 580).

It is unrealistic to try to show in a single paper that all of these broadly disjunctive strategies are unsuccessful. However, I think that a strong case can be made that the unified belief-based account sketched here should be preferred. For, there are general theoretical reasons to prefer a unified psychological account of the placebo effect to any heterogeneous disjunctive account, and there are specific problems with any disjunctive account that includes conditioning and belief-based accounts as distinct components.

First, consider the general explanatory virtues of unification. On the influential unificationist account of scientific explanation, explanation simply consists in providing a unified account of a variety of phenomena (e.g. Kitcher, 1989). So, if we adopted this general theory of explanation, we would have reason to prefer a unified account to a disjunctive one.
However, even those who reject the unificationist theory of explanation and endorse a competing causal theory of explanation accept that a particular kind of unification is explanatory. Namely, unification is explanatory when apparently disparate phenomena are “shown to be the result of a common set of mechanisms or causal relationships” (Woodward, 2009, §5.4). Other things being equal, an account of different instances of the placebo effect that shows them to result from a common psychological mechanism should be preferred to an account that treats those instances as products of many different kinds of psychological (or biochemical) mechanism.

What about an eliminative account, mentioned above, that eschews psychological mechanisms entirely? I think that Miller & Kaptchuk have the right kind of response to such proposals: “From a biological perspective, there are multiple placebo effects. It remains an open question whether there is any common psychological mechanism that explains such effects” (2008, p. 223). This paper attempts to answer this open question. Note that even if many different biochemical mechanisms are responsible for placebo effects (as outlined in Benedetti, 2008, p. 52), this is compatible with a single psychological explanation for placebo effects. The common psychological mechanism may be multiply realized by different biochemical pathways.

That being said, one need not rely on these general considerations to reject a heterogeneous psychological account of the placebo effect that includes conditioning and belief-based mechanisms as distinct components, for this particular disjunctive account faces specific problems, to which I now turn.

3.1. Unconscious Beliefs and the Inadequacy of Behaviorist Accounts of Conditioning

The main problem with the conditioning account is that it is unclear if it offers an explanation of instances of the placebo effect that is independent of cognitive or belief-based accounts; that it is, it is doubtful whether the conditioning account is a genuine theoretical
alternative to cognitive accounts. Most contemporary theories of conditioning hold that most conditioning processes in humans are themselves mediated by cognitive states like expectancies (e.g., Rescorla, 1988; Kirsch et al., 2004).⁴

Although these theories are not entirely uncontroversial, assuming something like them is correct, the difference between conditioning and expectancy accounts turns out to be unimportant. In effect, the conditioning account would collapse into the cognitive account as a special case.⁵ Of course, if this occurred, then we would lose conditioning as a distinct disjunct in a proposed heterogeneous account of placebo effects, and we could not appeal to conditioning to explain instances of the placebo effect that posed problems for belief-based accounts (such as those discussed in Section 2.2.). However, in order for a belief(expectancy account to cover such cases, it needs to hold that unconscious beliefs or expectancies play a crucial role in the etiology of some placebo effects. For example, it is highly implausible to claim that the effects on respiration and hormone levels discussed above are mediated by conscious, explicit expectancies about these autonomic, involuntary processes.

This is already a significant modification of the most common versions of the cognitive account, which have tended to assume that beliefs and expectancies are (perhaps by definition) conscious or explicit (e.g. Montgomery & Kirsch, 1997, p. 108) and that the conscious-unconscious distinction maps perfectly onto the expectation-conditioning distinction.⁶ However, even if these assumptions are dropped and one allows for unconscious or implicit beliefs, there are significant problems that face this liberalized version of the cognitive account. In Section 4.3, I identify several criteria that psychological states must satisfy in order to be classified as beliefs. I then argue that some states that mediate instances of the placebo effect do not meet these criteria. Before turning to these more general problems with belief-based accounts, I first want
to discuss an interesting puzzle that arises if the placebo effect is mediated by a belief with a certain kind of content: namely, that believing in the placebo response in one’s own case is necessarily irrational, if not paradoxical.

3.2. The Irrationality of Believing in Some Belief-Mediated Placebo Responses

In one’s own case, a plausible belief-based placebo hypothesis is: “I will get better solely because I believe I will get better.” But can one be justified in believing this hypothesis? If the hypothesis is true, in order to get better, I need to believe that I will get better. But on what grounds should I believe this? Suppose I believe that I will get better for some other reason. If I truly believe that I will get better because of this reason, then the placebo hypothesis cannot be true; for then I will get better not solely because I believe that I will. If I falsely believe that I will get better for this other reason, then either I will not get better or I will get better but not for the reason I think. In the former case, the placebo hypothesis is false, and in the latter case, my belief that I will get better is not justified, so if I am rational I shouldn’t hold it. Perhaps I should believe that I will get better simply because I will get better; maybe my belief is self-justifying! But, if the placebo hypothesis is true, one of the effects of my belief that I will get better is my getting better, and it is usually not rational to treat the effects of a belief as evidence for it. As Roger Squires puts it:

[A] medically well informed person cannot take advantage of the placebo effect without adopting unreasonable beliefs. There is no paradox here, however disappointing to the philosopher. Epistemic rationality may be maintained …, but at the price of persistent illness …$. That it is not always best to believe what is most likely to be true may raise no eyebrows in hospitals …$, but I find it strangely unsettling” (1994, p. 375)

Peter Cave (2001), in fact, argues that there is a paradox here which is similar to Moore’s paradox. Suppose I take a pill that has no pharmacological curative powers. I initially neither know nor believe the pill to be a placebo, but I know that I have taken the pill and believe that I
will be made better as a result. Suppose I come to entertain the placebo hypothesis: “I will get better solely because I believe the pill will make me better.” Can I coherently come to even believe this hypothesis? I cannot, for when I try to do so, my belief that the pill will make me better is undermined and hence so is any tentative belief in the placebo hypothesis itself. As Cave puts it, I cannot rationally believe the placebo hypothesis because of a “deviant dependency” between what is believed and the reason for what is believed:

When [I] entertain [the placebo hypothesis], upon placebo discovery, [I] may initially still accept [(r) that the pill will make me better], but, as reasoning progresses, r, if accepted at all, is seen to be accepted solely because of [(s) that I believe that the pill will make me better], yet s’s truth seemingly is secure because it is but the belief that r which [I] accept. [My] s might be thought sound and safe – for if a proposition is true and one believes it, things are going well – but s is reason-hungry, its hunger formerly being satisfied by [my] erroneous belief in the pill’s curative pharmaceutical properties. That latter belief is gone, r is sole contender as reason for s, if s is still to be held; yet r is no reason at all, being no more than that which s embeds. While it is true that one believes a proposition p true, if one believes that p – and the belief that p is made true by p’s truth – p cannot sensibly be cited as reason for holding the content of one’s belief that p is true, though it may sensibly figure in the explanation of why someone comes to believe that p. (2001, p. 144)

These consequences by themselves of course do not show that this particular belief-based account of the placebo response is mistaken. Perhaps we should simply learn to accept that it is not always best to believe what is most likely to be true. However, it seems that we should avoid this result if at all possible. If this account were true, it would be impossible to hold a rational belief that the placebo response will occur in one’s own case. Placebo researchers would never be able to benefit from their own work, and not for lack of trying! While not an outright contradiction, this account of the placebo response relegates it to the shadows and results in superstition being an essential part of the placebo response.

Of course, the problem of irrationality arises in these cases because a belief of a particular form is involved in the account, and this does not show that all expectancy accounts
raise the same problem. For instance, perhaps one gets better solely because one believes that
one’s doctor is highly skilled, that she cares about one’s health, and is enthusiastic about a
particular treatment. Since these beliefs do not embed the content that one will get better, they do
not raise the problems discussed above. However, there are also problems with belief-based
accounts in general.

3.3. The Cognitive States that Mediate Placebo Responses Do Not Possess Some Characteristic
Features of Beliefs

While some studies, such as Benedetti and colleagues’ (1998) work on respiration, may
be able to be accommodated by a cognitive account merely by appealing to unconscious beliefs,
others continue to pose problems. For example, the subjects in the hormone portion of (Benedetti
et al., 2003) who were told that their cortisol levels would increase when given a saline injection
(and hence presumably had a belief with that content, if they had any belief on the matter) still
had their cortisol levels decrease (because of the prior conditioning injections of sumatriptan).
Now, a proponent of the belief-based account could claim that the subjects also unconsiously
believed that their cortisol levels would decrease and that this belief is responsible for the
placebo response. Unconscious beliefs are certainly common. For example, while writing this
paper it has been true of me that I believe that the sun is a star, but I certainly haven’t been
consciously aware of that belief until now. Most of what I believe is like this: the experience of
having the belief occurrently is provoked when my attention is drawn to the content of the belief
(e.g. by asking what kind of heavenly body the sun is.) This indicates that even unconscious
beliefs have contents that are consciously accessible.?

Given this feature of beliefs, the attempt to construe the decrease in cortisol as being
brought about by an unconscious belief looks like special pleading. When prompted, subjects
who were told that their cortisol levels would increase would surely claim that they believed this.
If they also had a contradictory unconscious belief, one would expect that they would at least display some uncertainty or confusion, given that this latent inconsistency would have been brought to their attention.

In the examples involving “formal” features, it is not as clear that subjects do not have unconscious beliefs about the relative efficacy of different colors, numbers, prices, brands, or delivery methods of treatments. However, it is also not entirely clear that the subjects do have such unconscious beliefs. It seems that these cases could be equally well described by claiming that the subjects were in some cognitive state that represents information about these formal features and their associated comparative efficacy that, while not a belief itself, can give rise to a belief in appropriate conditions. If this construal is correct, then what goes on when subjects’ attention is drawn to these features is not analogous to what occurs when we ask them where they were born (while talking about another subject). We are not bringing to consciousness a belief that they had all along. Rather, we are setting off a process of belief-formation, in which a subdoxastic cognitive state with a certain content may play a crucial role in producing a belief with very similar content.8 (I call any cognitive state ‘subdoxastic’ if (at least some of) its content is not consciously accessible and is largely inferentially isolated from a subject’s network of beliefs (see next section).

Is there any reason to prefer one of these descriptions to the other? Well, in addition to being consciously accessible, beliefs are also inferentially promiscuous: “Provided with a suitable set of supplementary beliefs, almost any belief can play a role in the inference to any other” (Stich, 1978, p. 506). By contrast, states that bring about some placebo responses are largely inferentially isolated from the large body of integrated beliefs to which a subject has access.9 For example, compare the (putatively subdoxastic) state brought about by conditioning
that represents the information that one’s cortisol levels will decrease with the belief that one’s cortisol levels will decrease. Together with other beliefs, the latter can be used to infer just about any other belief: if the subject also believes that testosterone levels will decrease, she may infer that two hormones will decrease. Further, just about any other belief, b, can inferentially give rise to the belief that one’s cortisol levels will decrease together with, say, the belief that if b, then one’s cortisol levels will decrease. However, the state produced by conditioning can lead to only a limited number of beliefs, if any. For example, if a subject believes that her testosterone will decrease, she will not infer that two hormones will decrease (unless, of course, the subdoxastic state produces the conscious belief that cortisol will decrease). And, it is plausible that no, or at least very few, beliefs can inferentially lead to the state that is produced by conditioning (more on this below). Similar claims hold for the states involved in the formal features examples; they have very few potential inferential connections to other beliefs.

There is one more feature of beliefs that may not be possessed by the states that give rise to some placebo responses: some of these states seem to directly cause behavioral changes or other physiological effects (for example, in the immune responses discussed above). By contrast, beliefs cause behavior and physiological changes only in conjunction with other mental states, such as desires. I do not rely on this as an argument against expectancy-based accounts, however, because some researchers claim that expectancies can directly and immediately produce certain physiological results (see Stewart-Williams and Podd (2004, p. 329) for discussion). I suspect that ‘expectancy’ as used by many psychologists is not fine-grained enough to distinguish beliefs from the subdoxastic states that, according to my proposal, cause placebo responses.
4. A New Proposal: Aliefs Mediate Placebo Effects

Cognitive states with *consciously inaccessible* content which is *inferentially isolated* from the subject’s large network of beliefs (and which may *directly* cause behavior) are not beliefs. As noted above, I call such cognitive states *subdoxastic*. Subdoxastic states may have content that is either propositional or non-propositional and may not involve acceptance of that content (see note 12). (Stephen Stich uses the term ‘subdoxastic’ more narrowly to refer to “psychological states that play a role in the proximate causal history of beliefs, though they are not beliefs themselves” (1978, p. 499). According to my usage, a subdoxastic state may play a role in causing a belief, but it need not.)

The evidence presented in the previous section indicates that subdoxastic states play a role in the production of many placebo effects. In particular, the subdoxastic cognitive state called “alief,” which has recently been described and defended by Tamar Szabó (2008(a), 2008(b)), is a promising candidate for a unified account of the placebo effect. Gendler claims that alief is more primitive than belief or imagination and plays a far larger role in causing behavior than much of the recent philosophical literature would suggest. Paradigmatic aliefs are mental states with “associatively linked content that is representational, affective and behavioral, and that is activated – consciously or nonconsciously – by features of the subject’s internal or ambient environment” (2008(a), p. 642). Aliefs are so-called because they are “associative, automatic, and arational. As a class, aliefs are states that we share with nonhuman animals; they are developmentally and conceptually antecedent to other cognitive attitudes that the creature may go on to develop. And they are typically also affect-laden and action-generating” (ibid., p. 641). To take one of Gendler’s examples, when someone voluntarily walks out on the glass-bottomed Grand Canyon Skywalk and sees the Colorado River rushing 4,000 feet below her, she
believes that the walkway is structurally sound and that she is perfectly safe, but she has an alief
with roughly the content: “Really high up, long long way down. Not a safe place to be! Get
off!!” In this case, the contents of the alief and of the belief conflict, but it is the alief that
explains the person’s behavior and physiological responses: her trembling hands and racing
heart. Gendler suggests that it is also alief that drives behavior in cases where one’s alief and
belief are concordant (ibid., pp. 646 and 663).

I propose that, in general, aliefs are the subdoxastic states that play an essential role in the
production of placebo responses.11 For instance, when researchers tell their subjects that an
iodine mixture is a powerful painkiller they consciously activate an alief that has roughly the
content: “Pain reliever administered. Soothing and analgesic. Produce endogenous opioids
(among other physiological responses).”12 This verbal prompt also produces a belief that a pain
relieving ointment had been applied, but if Gendler’s suggestion is correct, it is the alief that is
responsible for the reduction in pain.

5. Benefits of the Alief-Based Account

Claiming that placebo responses and effects are caused by aliefs has several benefits. It
handles the problem cases discussed above better than existing accounts, it allows us to improve
and refine our account of aliefs, and it has the potential to reveal interesting connections to other
psychological phenomena.

5.1. Addressing the Problems Facing Existing Accounts

Belief-based accounts fare fairly well with the examples discussed in 2.1., but they have
trouble with the cases in 2.2 because the states that mediate these effects do not satisfy two
criteria on beliefs; they do not have consciously accessible contents, and they are inferentially
isolated. On the other hand, conditioning accounts can explain the latter cases, but they fare
poorly with the former because these placebo effects are clearly mediated by internal states with some kind of content. However, the alief account is able to handle all of these cases nicely.

Consider the portion of Montgomery & Kirsch’s (1997) study discussed in 2.1. For the uninformed group, the shocks during the manipulation trials produced an alief with roughly the content: “Pain reliever administered. Soothing and analgesic. Produce endogenous opioids (among other physiological responses).” The contextual cues were unambiguous, and their alief systems smoothly produced and activated the relevant affective and physiological routines. However, for the informed group, the shocks in the manipulation trials occurred in a different context. Informed subjects had been previously told that shocks paired with application of the placebo were at a lower intensity than shocks given without the placebo. The subsequent inputs into the alief system were ambiguous since subjects were uncertain about how much of the pain reduction was due to the Trivaricane and how much was due to the reduction in shock intensity. Since their alief system was given ambiguous cues, it is plausible that it produced an alief associated with placebo application with less intense affective and physiological components than the aliefs in the uninformed group.

The investigators’ verbal suggestions probably also produced different beliefs in the two groups. Those in the uninformed group may have believed that their pain was reduced solely because of the application of the ointment, while those in the informed group may have lacked this belief. The informed subjects may have believed that their pain was reduced partially because of the ointment and partially because of the change in stimulus intensity. However, if subjects had these beliefs, it is less plausible that they directly affected pain levels than that they contributed to changes in the subjects’ aliefs which brought about the relevant changes in perceived pain.
Turning to the pain portion of Benedetti’s (2003) study, it is plausible that two aliefs were activated in the subjects. First, there was a belief-concordant alief consciously induced by the investigators verbal suggestion with the rough content: “Pain promoter administered. Aggravating and hyperalgesic. Suppress endogenous opioids (among other physiological responses).” Second, there was a belief-discordant alief unconsciously produced by the prior administration of the pain-reliever, ketorolac, with roughly opposite content. I think there are two plausible explanations for why the former alief “wins out” and results in increased pain. First, the former alief has roughly the same content as the belief that was also produced in the subjects by the verbal suggestion, and this belief may help facilitate activation of the alief’s content (perhaps together with a desire). (There may be general, theoretical reasons to think that belief-concordant aliefs will dominate belief-discordant aliefs when both are present.) Second, setting aside the relevant belief, paradigmatic pains are consciously experienced psychological states. Thus, given the modular nature of aliefs, consciously activated aliefs are more likely to cause changes in consciously experienced pain than unconsciously activated aliefs (when both are present) since the former is likely to be a component in a (set of) module(s) governing conscious experience.

One objection to both of these explanations is that they merely reproduce the expectancy-based and conditioning accounts with different terminology. It is true that these explanations are superficially similar in some respects to previous accounts, but there are important differences. Previous accounts either saw different placebo effects as the result of fundamentally different psychological processes (see, e.g., note 5), or attempted to assimilate conditioning to belief-based processes. My proposal does neither, for the reasons given above. Rather, it invokes a kind of subdoxastic state that underlies, and is developmentally prior to, both conditioning and traditional belief/desire psychology. Aliefs are a distinct kind of cognitive state that provide a
“common currency” that can mediate both conscious and non-conscious processes. As Gendler puts it, to a first approximation, to have an alief is “to have an innate or habitual propensity to respond to an apparent stimulus in a particular way” (2008(b), p. 557, italics added). For example, the hormone and respiration studies involve innate aliefs that are engendered by prior exposure to the relevant drugs, while the responses to some formal features of treatment are habitual aliefs produced via prior interaction with the socio-cultural context. (For this reason, I suspect that Moerman’s “meaning response”: “the psychological and physiological effects of meaning in the treatment of illness” (2002, p. 14), is too coarse as a theoretical construct. It wrongly assimilates complicated belief-based, reason-responsive phenomena, such as psychotherapy (ibid., p. 94ff.) to alief-mediated phenomena like responses to formal features and to a doctor’s enthusiasm for a drug. This example may illustrate that unification per se is not explanatory, while unification that identifies common causal mechanisms is.)

Benedetti and colleagues’ respiration (1998) and hormone (2003) results can be explained more briefly. When Benedetti gave his subjects a saline injection, after having previously given them buprenorphine injections, he unconsciously activated an alief with roughly the content: “Same drug given as before. Produce similar affective and hormonal states.” This explains why subjects were neither aware of the content of this alief nor of its effects (mild respiratory depression). Further, this example illustrates the modularity and inferential isolation of aliefs, since those that produce respiratory depression were independent of the aliefs that resulted in analgesia (in different subjects). Similarly, when Benedetti injected his patients with saline, after having previously given them two injections of sumatriptan, he unconsciously activated an alief that has roughly the same content as above. When he told them that their cortisol levels will increase, which contradicts the content of the “conditioned” alief, they may
have believed this, but their alief system was largely in control of the subsequent physiological changes in their bodies.

Aliefs also provide a more natural treatment of placebo effects involving formal features of treatment. When subjects in a particular cultural context are exposed to a formal feature of a treatment, like the fact that it is in injection form, rather than a pill, an alief is consciously or unconsciously activated in them whose content will vary depending on the cultural associations of that formal feature. For example, when an American is given an injection, an alief is activated with roughly the content: “Injection given. Powerful medicine. Heighten certain physiological responses.”

Finally, aliefs provide a tidy solution to the irrationality problem from Section 3.2. Prior life experience results in an intricate, largely implicit set of affective and behavioral or physiological associations involving these features. When these features appear in a treatment paradigm, these associations are activated automatically (see the references in note 7). Subjects need not have explicit or implicit beliefs about these features, which would only produce behavioral or physiological changes in conjunction with other mental states. Similarly, we can avoid the puzzles about irrationality that face belief-based accounts. In order for the placebo response to occur, one need not believe that one will get better (or that a pill will make one better). Rather, one only needs to alieve a state with something like this content. Since alief is arational, there need be no irrationality in a medically well-informed person benefiting from the placebo response.¹³

Note that these last two benefits hold even if one was unconvinced of the benefits of a unified psychological account of placebo effects. Thus, even if someone endorsed a disjunctive
account, at the very least this paper has shown that a subdoxastic alief-based model should be on
the table as one among many psychological mechanisms for placebo effects.

5.2. Elaborating and Refining the Account of Aliefs

Gendler emphasizes that it is only *paradigmatic* aliefs whose content can be
characterized as representational, affective, and behavioral. As she writes:

> Perhaps there are cases where the activation occurs at a sufficiently low level to
render the notion of representation inapplicable. Perhaps there are states that lack
an obvious affective ingredient, or that do not include the clear activation of a
motor routine, but that nonetheless sufficiently resemble our paradigm cases that
we want to count them as aliefs. Perhaps there are cases where the most
noticeable associations are not easily subsumed under the three categories offered
– cases that primarily involve the heightening or dampening of certain sorts of
attention, or the heightening or dampening of certain perceptual sensitivities.
(2008(a), p. 644)

If aliefs play the role in the placebo response I have suggested, then some of them may not be
accurately described as fully representational (e.g. those produced by conditioning trials).

Further, the associations of such aliefs are often not best described as behavioral; in the cases
described above they involve heightening or dampening of certain physiological or
psychological responses.¹⁴

In addition, although Gendler does not emphasize it, the inferential isolation of aliefs
suggests that they are components of modular subsystems (e.g., pain, hormonal, etc.) whose
processing does not have significant informational connections to other subsystems. This would
explain why aliefs tend to be cognitively impenetrable; they persist even in the face of
contradictory beliefs.

Finally, my proposal suggests promising lines of research into the neurological substrates
underlying some aliefs. As several imaging studies have shown (e.g. Wager et al., 2004; Petrovic
et al., 2005), during the production of placebo responses both affective areas (e.g. anterior
cingulate cortex and amygdala) and modulatory or representational areas (e.g. the dorsolateral prefrontal cortex and lateral orbitofrontal cortex) of the brain are activated. This is exactly what should be expected if a state that has content with representational, affective, and behavioral/physiological components, like alief, is involved in the placebo response.

6. Concluding Remarks

I have argued that appealing to aliefs elucidates many puzzling features of placebo effects and holds promise as a unified account of such effects. Assuming that some scientific explanations proceed by showing that apparently disparate phenomena are all the result of a common causal mechanism, support for the alief-based account would be strengthened by showing that it also covers other perplexing psychological phenomena. I close by mentioning an intriguing test case: mirror box therapy for phantom limb pain and paralysis. In such therapy, patients move the amputated limb while viewing mirror images of their intact limb performing corresponding movements. One “tricks the brain” into perceiving movement in the phantom when the intact limb is moved, in effect, giving subjects a “placebo arm” (see, e.g., Chan et al., 2007; Ramachandran & Rogers-Ramachandran, 1996). Presumably, patients who receive mirror treatment do not believe that their phantom limbs are miraculously reattached and moving again. However, it is plausible that they have an alief with something like this content. Perhaps this alief resets certain maladaptive patterns of neural firing, correcting inconsistencies between motor signals to the phantom limb and a lack of congruent sensory feedback.

The prospect of this kind of explanatory unification would have been lost if one had abandoned the attempt to give a unified psychological account of the placebo effect and instead claimed that it should be sundered into unrelated psychological or biochemical processes. If this paper is correct, aliefs can provide such a unified psychological account. The alief-based
account thus allows us to avoid one of the dangers in letting biochemical and neurological research completely dominate an interdisciplinary research agenda.

Acknowledgements

I would like to thank Tamar Szabó Gendler and Esther Sternberg whose talks at “The Study of the Human Self” conference at William & Mary in fall 2008 prompted me to try to explain the placebo effect in terms of aliefs. Thanks also to Paul S. Davies for helpful feedback on an early draft and to Stephen Crowley for comments on a shorter version of this paper, which I presented at the 2010 Society for Philosophy and Psychology meeting in Portland, OR. I am grateful for financial support from a Faculty Summer Research Grant from the College of William & Mary.
Notes

1 Some authors have regimented terminology as follows (e.g. Hoffman et al., 2005). The placebo response occurs in a given individual and is the difference between the untreated, natural history condition and the individual’s condition after administration of a placebo (i.e. after interaction with elements in the psychosocial context). By contrast, the term ‘placebo effect’ refers to the average response of members of a group to placebo manipulation (i.e. the average placebo response for a particular group of subjects). Not all authors adopt this usage, and I occasionally depart from it in this paper.

2 Gendler does not claim to have discovered a new category of mental states. Rather, she claims merely “to have noticed a certain commonality across some lines of thought that might otherwise have appeared disparate” (2008(b), p. 557 n.5). In this paper, I argue that this commonality extends to the placebo effect, as well.

3 There is compelling brain imaging work that supports the view that the placebo effect involves an actual reduction in the intensity of experienced pain and not merely in a reduction in the amount of pain that is reported (e.g. Wager et al., 2004).

4 As has been repeatedly pointed out in the placebo literature (e.g., Stewart-Williams and Podd, 2004; Hoffman et al., 2005; Voudouris, 1990), the conditioning and expectancy accounts are not mutually exclusive, since response expectancies can be developed through conditioning. However, even Kirsch and colleagues admit that some conditioning phenomena – including conditioned immunosuppression, conditioned taste aversion, and subliminal conditioned stimuli – do not appear to be cognitively mediated (at least by conscious beliefs or expectancies – more on this below).
An anonymous referee pointed out that someone could hold that conditioning is mediated by representational states other than beliefs or desires, and the resulting view would not collapse into a cognitivist view like those discussed in Section 1.2. I agree, and I think that aliefs should be preferred to beliefs in any such account, for the reasons given in Sections 3.2 and 3.3. See also Section 5.1.

“[P]lacebo responses are mediated by conditioning when unconscious physiological functions such as hormonal secretion are involved, whereas they are mediated by expectation when conscious physiological processes such as pain and motor performance come into play, even though a conditioning procedure is performed” (Benedetti et al., 2003, p. 4315).

Perhaps there are some unconscious beliefs that cannot be brought to awareness, or at least not readily, e.g. those that are repressed. However, in these cases, there is plausibly some mechanism that intervenes to block conscious access (cf. Stich (1978, p. 505).

If this latter description is correct, as I argue below, then exposure to the formal features of a treatment unconsciously “primes” subjects in much the same way that performing scrambled sentence tasks does. Such priming results in *automatic*, unconscious activation of the associated affective and behavioral routines. See, e.g. Bargh et al. (1996), Gendler (2008a, p. 656-661).

Following Stich, I assume that inference is not by definition a relation solely between beliefs (1978, pp. 507, 511-517).

Note that Gendler uses the term ‘content’ in a somewhat idiosyncratic, general way that may include affective states and behavioral dispositions (2008a, p. 635 n.4). As I discuss below, the contents of aliefs involved in the placebo response also involve physiological dispositions that one may be reluctant to call behavioral. Also, even if an alief is consciously activated, usually not all of its content is consciously accessible.
One may also want to allow a role for the subdoxastic analog of desire, which Gendler dubs “cesire” (2008a, p. 642 n.17), for similar reasons to those given by Price & Fields (1997), mentioned above.

These descriptions of the representational, affective, and physiological components of the alief’s content are stilted and artificial, and for good reason: I suspect that the contents of the relevant aliefs are non-propositional and resist description in natural language.

In fact, there is some intriguing, although almost anecdotal, evidence that suggests that one can experience a placebo response even if one is convinced (Park & Covi, 1965), or has some reason to believe (Bergmann et al., 1994), that one is receiving merely a placebo treatment. If such cases are corroborated, then they provide another example of alieving a content (the pill will make me better) without accepting it, and thus not believing it (see Gendler, 2008, p. 648-651).

However, the aliefs involved in some placebo responses do have behavioral content. For example, in a double-blind study, McRae et al. (2004) performed sham surgery for some patients with Parkinson’s. (I.e. brain surgery was performed on all subjects but the half in the placebo arm did not actually receive transplants of embryonic dopamine neurons.) Even the subjects who received the placebo surgery showed significant improvement in physical, emotional, and social functioning, even after one year.
References


Kitcher, P. (1989) Scientific explanation and the causal structure of the world. In P. Kitcher and
W. Salmon (Eds.), *Scientific Explanation*. (pp. 410-505). Minneapolis: University of
Minnesota Press.

McRae, C., Cherin, E., Yamazaki, T.G., Diem, G., Vo, A.H., Russell, D., Ellgring, J.H., Fahn, S.,
perceived treatment on quality of life and medical outcomes in a double-blind placebo

Miller, F.G. & Kaptchuk, T.J. (2008) The power of context: Reconceptualizing the placebo

UP.

107-113.

45.

in emotion processing: Induced expectations of anxiety relief activate a generalized


