The Double-Blind Randomized Controlled Trial as the Gold Standard in Psychedelic Research: Neither Feasible nor Desirable

Daniel Villiger

This article has been accepted for publication in the
Kennedy Institute of Ethics Journal, 2024, following peer review.

Abstract

Double-blind, randomized, controlled trials (DB-RCT), if well-designed and -conducted, are widely considered the gold standard in medical research for purposes of establishing causal efficacy. Their logic is compelling: by balancing out all confounding variables through the research design, DB-RCTs are thought to reveal whether a proposed treatment—by virtue of its characteristic constituents—*causes* therapeutic effects. Many studies on psychedelic-assisted therapy (PAT) follow this ostensible gold standard and use a DB-RCT design. However, several authors have already noted that conducting psychedelic DB-RCTs is particularly challenging: due to the psychoactive effects of psychedelics, participant awareness of condition assignment is likely; this awareness may then interact with response expectancy and experimenter behavior, introducing systematic bias into the trial. For this reason, these authors have suggested ways to rescue DB-RCTs for PAT. The present paper takes a different direction. It argues that we should abandon the DB-RCT design as the assumed gold standard in PAT research, because its logic is largely undermined by the intervention(s) in question, and the design in its standard form neglects potentially important aspects of PAT (i.e., extra-pharmacological factors and their interaction(s) with the psychedelic). Doing so opens the door to a more holistic study of PAT, in which DB-RCTs are still useful for certain ends, but no longer considered to produce results that are per se superior to those of other research designs, but complementary.

1 Introduction

After decades of obscurity, psychedelic-assisted therapy (PAT) has once again become a much-researched topic in psychiatry.[1] In the 1950s and 1960s more than 1000 papers were published on psychedelics, including clinical studies with around 40,000 patients overall (Grinspoon and Bakalar 1997). However, this first flowering of psychedelic research came to an abrupt halt after the United States Drug Enforcement Agency classified psychedelics as Schedule I substances in 1970. It is important to note that this quasi-blanket ban was enacted despite previous evidence showing that psychedelics are largely non-addictive and PAT could have positive medical uses (Johnson et al. 2018). It took more than 30 years before a new PAT study was conducted: From 2001 to 2004, Moreno et al. (2006) applied PAT to nine patients suffering from obsessive-compulsive disorder with positive results.[2] While this could be seen as marking the start of a new era of clinical psychedelic research, the beginning of this new era was slow. The next study on PAT was published in 2011 and then in 2014 (Gasser et al. 2014; Grob et al. 2011). But from then on, the number of studies steadily increased, reaching over one hundred ongoing studies on PAT as of November 2024 (ClinicalTrials 2024).

One main reason why PAT has once again become a much-researched topic is that modern PAT studies have so far produced what are widely considered to be promising results. Systematic reviews of such studies suggest that PAT can, and often does, have substantial therapeutic effects and that it can be performed relatively safely (e.g., Bender and Hellerstein 2022; Maia, Beaussant, and Garcia 2022; Romeo et al. 2024), notwithstanding certain risks (cf. Smith and Sisti 2021). In addition, there is some evidence of efficacy for PAT in relation to various mental disorders, including (treatment-resistant) depression (e.g., Goodwin, Aaronson, et al. 2022; Rotz et al. 2023), anxiety with and without a life-threatening illness (e.g., Holze et al. 2023), alcohol and smoking cessation (e.g., Bogenschutz et al. 2015; Johnson, Garcia-Romeu, and Griffiths 2017), obsessive-compulsive disorder (e.g., Moreno et al. 2006), and anorexia nervosa (e.g., Peck et al. 2023).

---

[1] This paper focuses on PAT with classic serotonergic psychedelics, including lysergic acid diethylamide (LSD), psilocybin, and N,N-dimethyltryptamine (DMT). 3,4-methylenedioxymethamphetamine (MDMA) and ketamine are sometimes classified as psychedelics too, so some also refer to a therapy supported by one of these substances as PAT. Moreover, there is also a lot of research on therapy supported by one of these substances, including phase 3 studies (Loo et al. 2023; Mitchell et al. 2023). However, these substances and the therapies associated with them will not be discussed in this paper.

[2] Around the same time as Moreno et al., Griffiths et al. (2006) conducted empirical psychedelic research with healthy participants to study psychedelic-induced mystical experiences.

However, the validity of modern PAT studies' apparently striking positive results has been questioned. Muthukumaraswamy et al. (2021) argue that treatment effect sizes in psychedelic double-blind randomized controlled trials (DB-RCT) are likely overestimated due to condition anonymization (so-called "blinding") problems and response expectancy. Specifically, the psychoactive effects of psychedelics are often too pronounced to keep participants unaware of whether they are in the treatment group or in the control group. In addition, positive media reports about PAT have created a kind of psychedelic hype—sometimes also referred to as the "Pollan effect" in reference to Michael Pollan's (2018) very successful book and Netflix series (2022), both of which put psychedelics in a largely positive light—resulting in highly optimistic expectations about the efficacy of PAT (Yaden, Potash, and Griffiths 2022). In combination with unblinding, such expectations can confound PAT trials in a twofold way: On the one hand, the realization that one is in the treatment group may have a positive effect on treatment outcomes through response expectancy effects. On the other hand, the realization that one is in the control group may have a negative effect on treatment outcomes due to disappointment in being in the control group. Both confounds could increase the difference in treatment outcomes between treatment group and control group, resulting in overestimated effect sizes.

The issue of blinding and response expectancy has become an increasingly discussed topic in the PAT literature (Aday et al. 2022; Aday, Carhart-Harris, and Woolley 2023; Butler, Jelen, and Rucker 2022; Muthukumaraswamy 2023; Nayak, Bradley, et al. 2023; Schenberg 2021; Szigeti and Heifets 2024; Wen et al. 2024). The main approach to the issue is how to rescue DB-RCTs in psychedelic research, either by introducing new study designs or new causal inferential methods that can ultimately control for the confounding variables. This paper will take a different approach by questioning the basic assumption that DB-RCT should be the gold standard in PAT research to begin with. It will argue that the main goal of DB-RCTs cannot reliably, or perhaps even possibly, be achieved in PAT research and that always striving to adhere to this goal would neglect scientifically and medically important aspects of PAT. While this does not imply that DB-RCTs in PAT research are useless and should no longer be conducted, it does suggest that their presumed superiority compared to other research designs is faulty and should be rejected. The paper will discuss what sorts of research questions DB-RCTs still help to answer, and what other types of research designs—with their own limitations—could address questions for which DB-RCTs are unsuitable. Ultimately, combining results from these different research designs allows us to gain a more complete picture of PAT.

The argument of this paper is therefore not against the relevance of DB-RCTs in psychedelic research, as these designs remain useful for answering certain questions. It is against the premise that DB-RCTs should be considered the gold standard in psychedelic research,

implying that results from other research designs are inherently inferior to those from DB-RCTs. Introducing such a hierarchy between the DB-RCT design and other research designs is not justified in psychedelic research: On the one hand, the problems of blinding and response expectancy largely undermine the logic of DB-RCTs in psychedelic research, and any attempts to address these issues lead to other challenges. On the other hand, the (standard) DB-RCT design neglects potentially important aspects of PAT, namely the effect of the psychotherapeutic embedding and its interaction with the psychedelic. Therefore, it is only when psychedelic research makes use of a variety of research designs that it can balance the limitations of one design with the strengths of another. As will be shown, open-label randomized wait-list controlled trials and pragmatic clinical trials are such complementary research designs to the DB-RCT design.

The paper is structured as follows: Section 2 introduces the logic of the DB-RCT design and why it is seen superior to other research designs (assuming high-quality executions in either case). Section 3 analyzes the problems that psychedelic research poses to DB-RCTs. Section 4 discusses approaches proposed in the existing literatures that try to (partly) solve these problems, and why they ultimately fail. Section 5 examines why standard DB-RCTs likely neglect aspects of PAT that many consider constitutive of, or in any case necessary for, the treatment. Section 6 proposes how PAT should be studied, by analyzing which research designs are useful for which research questions.

2 The Logic of Double-Blind Randomized Controlled Trials

The DB-RCT design attempts to bring the ideal of scientific rigor to clinical trials in an effort to achieve an aspirational objectivity most closely associated with the laboratory model. One key aspect of this model is its focus on establishing internal validity (i.e., a high degree of confidence that the causal relationship being tested is trustworthy and not confounded). Due to this focus, the DB-RCT is widely considered to be the most reliable method in medical research for "representing things as they really are" (Rorty 1979, 334), in the sense of eliminating all alternative explanations, and has therefore become the gold standard in medicine (Kaptchuk 2001). It is interesting to note that the rise of the DB-RCT coincided with the first wave of psychedelic research in the 1950s and 60s described in the introduction, with the two developments affecting each other. On the one hand, although psychedelic researchers were among the first to implement the DB-RCT design in psychiatry, they strongly emphasized its limitations in this field (Dyck 2008). On the other hand, the difficulty of blinding psychedelic trials was

already criticized in the first wave of psychedelic research and was a major reason why certain promising clinical findings—for example, those obtained with LSD in people suffering from alcoholism—were distrusted (Schenberg 2021). Thus, the growing acceptance of the DB-RCT design as the gold standard has certainly not facilitated the flourishing of psychedelic research. Quite the opposite, the new gold standard has called into question the scientific objectivity of psychedelic studies and, consequently, the validity of the results found in such studies. But what exactly are the problems that the DB-RCT design is assumed to solve better than any other research design?

On dominant models of scientific validity, the validity of a treatment is not determined by whether it has therapeutic effects, but by whether the characteristic *constituents of the treatment* cause therapeutic effects; the term characteristic constituents, introduced by Grünbaum (1981), is used in placebo research (e.g., Wampold et al. 2005; Locher, Gaab, and Blease 2018) and refers to those constituents of a treatment that are theorized to produce its therapeutic effects. Thus, the scientific evaluation of a treatment must be able to exclude or control for confounding variables that are not part of the characteristic constituents of the treatment but may nevertheless influence therapeutic effects. To take a famous example, in 1784, a commission including Benjamin Franklin showed that so-called animal magnetism had no real effect: Under appropriately blinded conditions, therapeutic differences disappeared between individuals exposed to animal magnetism and those who were only told they were exposed to it. The commission concluded that illusions of the mind and subtle interactions between doctors and patients, as well as between patients and other patients, were responsible for the effects of animal magnetism (Gaab, Locher, and Trachsel 2021).

The total effect of the confounding variables is called the placebo response and includes factors such as placebo effects (i.e., effects due to belief in treatment), Hawthorne effects (i.e., effects due to being observed in a trial), spontaneous remission, and regression to the mean (Blease, Bernstein, and Locher 2020). DB-RCTs control for placebo response as follows: (1) They include, alongside the treatment group, a placebo control group: i.e., a group of individuals each of whom also receives the treatment but without its characteristic constituents; (2) they include a random assignment of participants to either group; (3) they include a blinding of participants and researchers as to which group participants are in. This way, bias introduced by the placebo response should be equally distributed across the treatment group and the placebo group. As a result, any difference between the treatment group and the placebo group can be attributed to the characteristic constituents of the treatment.

This logic behind the DB-RCT design is very compelling, and in fact it was the intuitive attractiveness of the design rather than empirical evidence that drove its rise to becoming the

perceived gold standard in medicine (Kaptchuk 1998). Nevertheless, it is not without its own problems. First, its focus on internal validity comes at the cost of external validity (i.e., the extent to which results from a study can be generalized to other contexts). In the context of normal treatment, both the patient and the clinician know that the former receives a real treatment and do not have to wonder about that. The epistemic uncertainty present in DB-RCTs can affect both the patient and the clinician (e.g., altered expectations of therapeutic success, stronger focus on changes to detect which group the patient is in) (Kaptchuk 2001). Second, in reference to internal validity, DB-RCTs only work flawlessly when therapeutic effects and possible side-effects are not easily noticeable by the unassisted senses.[3] For example, hypertension is a condition that is usually only detected by measuring blood pressure. The same is true for the successful pharmacological treatment of hypertension. This provides ideal conditions for DB-RCTs, because there is nothing that could unblind patients and clinicians during the study except measuring blood pressure. Now, as soon as therapeutic effects or possible side-effects are noticeable, unblinding is no longer guaranteed. On the one hand, patients receiving the real treatment may infer that they are in the treatment group by noticing side-effects (so-called malicious unblinding) or that their symptoms improve (so-called benign unblinding) (Howick 2011). On the other hand, patients receiving the placebo may notice that there are no side-effects or that their symptoms do not improve and infer that they are in the placebo group. Both cases of unblinding can lead to placebo or nocebo effects that are specific to one of the two groups, reintroducing systematic bias.

Obviously, the effects of psychedelics are far from being not easily noticeable by the unassisted senses and should therefore pose a challenge to the conduct of DB-RCTs in PAT research and thus to the DB-RCT design as the undisputed gold standard.[4] Let us examine this presumed challenge more closely.

3 Double-Blind Randomized Controlled Trials with Psychedelics

The phenomenology of a psychedelic experience includes features such as altered perception of time and space (Wackermann et al. 2008), (geometric) optical illusions (Aday et al. 2021), vivid autobiographical recollection (Grof 2008), paranoid and delusional thinking (Carhart-

[3] DB-RCTs also work if the characteristic constituents of the treatment are not therapeutic, in case of which there are no differences between the treatment group and the control group.

[4] Note that while this is true for the macrodose usage of psychedelics (which is standard in PAT trials), it is much less so for the microdose usage of psychedelics.

Harris et al. 2016), ego dissolution (Nour et al. 2016), near-death-like experiences (Timmermann et al. 2018), and more. This suggests that in a psychedelic DB-RCT, it should be clear to patients 1-2 hours after administration whether they have received a psychedelic or a placebo (at least if it is an inert placebo and patients have some knowledge of the phenomenology of a psychedelic experience).

A recent systematic review of 48 randomized controlled PAT trials published between 1963 and 2023 confirms this (Wen et al. 2024). Of the 48 trials, 34 were double-blind (or at least attempted to be) and 35 included inert placebos. 14 studies reported patient blinding assessments and 7 studies reported clinician blinding assessments. Blinding assessments show that virtually all participants and clinicians were able to correctly identify which group participants were in. So, it can be concluded that blinding in PAT trials has been unsuccessful. Another systematic review that examines both clinical and non-clinical psychedelic studies reaches a similar conclusion (Nayak, Bradley, et al. 2023).

There is some evidence that placebo trips are possible. In a study by Olson et al. (2020), 61% of participants reported some effect even though they were given a placebo. Some of the participants even showed psychoactive effects on a scale typically seen in moderate or high doses of psilocybin. It should be noted, however, that these results come from a group experiment in which participants thought they were getting a psychedelic for sure, even though everyone was getting a placebo, and in which several confederates mimicked a psychedelic trip. Such features are obviously not present in PAT studies, which is why placebo trips are much less likely in such studies, increasing the likelihood of unblinding.

The unsuccessful blinding in PAT trials opens the door for condition-specific response expectancy and concomitant systematic bias. It can be assumed that the greater the difference between condition-specific expectations, the greater the potential for systematic bias. This in turn implies that the more PAT and its psychoactive effects are associated with therapeutic expectations while placebo treatment is not, the more response expectancy undermines the DB-RCT design.[5] Now, patients do not enter PAT trials as blank slates but have pre-formed expectations about the efficacy of PAT. These expectations are likely influenced by the hype surrounding the so-called psychedelic renaissance (Yaden, Potash, and Griffiths 2022): in recent years, a flood of research papers, popular publications, podcasts, media articles, and social media personalities have heralded psychedelics as "wonder drugs," generating considerable attention and enthusiasm for the psychedelic renaissance. Such positive expectations about the

[5] For example, if patients in a DB-RCT do not expect a treatment to be therapeutic anyway, unblinding is less of a problem because there are no group-specific expectancy responses.

efficacy of PAT may not only lead to corresponding response expectancy, but also to detrimental effects in the control group: the realization that one is not getting the "wonder drug" but rather a placebo may lead to disappointment and demoralization, even despair, potentially worsening outcomes and thereby increasing systematic bias.

So far, there is only one psychedelic DB-RCT that analyzed expectancy responses (Carhart-Harris et al. 2021). This trial compared psilocybin with escitalopram (an SSRI antidepressant) in patients with depression: over a 6-week period, patients received either "two separate doses of 25 mg psilocybin 3 weeks apart plus 6 weeks of daily placebo (psilocybin group) or two separate doses of 1 mg psilocybin 3 weeks apart plus 6 weeks of daily oral escitalopram (escitalopram group)" (1402). First, to the astonishment of psychedelic enthusiasts, the study found no significant differences between the two groups on the predetermined primary outcome (see also Nayak, Bari, et al. 2023). Second, a recent study that analyzed response expectancy in this trial found that: (1) patients had significantly greater therapeutic expectancies for psilocybin than for escitalopram; (2) therapeutic expectancies for escitalopram were associated with improved therapeutic outcomes for escitalopram; and (3) therapeutic expectancies for psilocybin were not predictive of response to psilocybin (Szigeti et al. 2024). The last finding is surprising because the "set" (i.e., the person's expectations, physical and mental state) is considered central to the effects of a psychedelic experience (Hartogsohn 2016; 2020). It is important to note, however, that while the data does not confirm expectancy effects for psilocybin, equivalence testing shows that they cannot be ruled out either, emphasizing the need for further research on this topic (Szigeti and Heifets 2024).

Beyond psychedelic DB-RCTs, there is one psychedelic study that examined expectancy effects regarding personality change after ayahuasca use in a shamanic setting (Weiss et al. 2021). Using the five-factor model of personality, the study found that expecting favorable changes in neuroticism, extraversion, and conscientiousness correlated with greater corresponding changes in these dimensions after ayahuasca use. This provides some preliminary evidence for psychedelic response expectancy.

In conclusion, most psychedelic DB-RCTs do not evaluate the success of blinding, and those that do find almost all participants and clinicians unblinded, suggesting that blinding is most likely a hopeless endeavor in PAT research. And while the impact of unblinding on systematic bias via response expectancy may be less severe than suspected (cf. Szigeti et al. 2024), it ultimately remains unclear. So, it is safe to say that the logic of DB-RCTs is in danger of

being undermined in PAT research and therefore needs to be secured if one wants to defend the DB-RCT design as the gold standard.[6] Can this be done?

4 Saving Psychedelic Double-Blind Randomized Controlled Trials

There are three general (non-exclusive) approaches to secure the logic of DB-RCTs in PAT research. The first is to attempt to maintain blinding by having an active control group that experiences similar effects as the treatment group.[7] For example, some ayahuasca studies use a placebo designed to mimic some of the properties of ayahuasca, including color, taste, and smell (Dos Santos et al. 2021; Rocha et al. 2021), and in the case of one study even gastrointestinal discomfort (Palhano-Fontes et al. 2019). While in two of these studies 25% and 35% of the participants misclassified the placebo as ayahuasca, respectively, in one study there was no misclassification.[8] Based on these findings, Wen et al. (2024) argue that studies using LSD or psilocybin should also implement active placebos that mimic some of the common physiological effects of psychedelics. However, the success of such implementation is questionable. First, while ayahuasca is known for its unique color, taste, and smell, and the gastrointestinal discomfort it causes, psilocybin and LSD do not have such non-psychoactive distinguishing features. Second, while some participants in these studies mistook the placebo for ayahuasca, blinding was still largely unsuccessful (i.e., a far greater percentage of participants correctly identified what condition they were in than would be expected by chance). Third, a psilocybin study that used niacin as an active placebo—a substance that causes a mild physiological reaction without altering the psychological state and should therefore mimic some of the physiological effects of psilocybin—almost completely failed to maintain blinding (Grob et al. 2011).

A common practice in PAT trials is to use a lower dose of the psychedelic as an active placebo (Carhart-Harris et al. 2021; Gasser et al. 2014; Goodwin, Aaronson, et al. 2022). Doing

---

[6] Note that PAT is not the only medical intervention that poses a challenge for the DB-RCT design. For example, somewhat similar issues have been discussed in the context of research on mindfulness and meditation (e.g., Davidson and Kaszniak 2015; Gupta and Dhawan 2022) and psychotherapy (e.g., Wampold and Imel 2015).

[7] Some researchers have suggested that this should be combined with recruiting only psychedelic-naïve participants, as "deception should be easier to ensure in a sample of psychedelic-naïve participants" (Hovmand et al. 2023, 656). It should be noted, however, that there is currently no empirical evidence to support this. Furthermore, doing so leads to two potential problems: First, including only psychedelic-naïve patients introduces a sampling bias. Second, attempting to keep psychedelic-naïve patients in the dark about the psychoactive effects of psychedelics and the phenomenology of a psychedelic experience to ensure blinding likely interferes with the clinician's duty to disclose relevant information during the informed consent process (cf. Villiger 2024b; 2024a)

[8] Note that in all three studies participants were naïve to the use of ayahuasca.

so has two advantages: (1) Experimenters can tell patients that they will certainly receive the psychedelic, adding that there are several possible doses. This should reduce negative expectations associated with the possibility of being in the placebo group; (2) The control dose can produce some of the physiological and psychoactive effects of the treatment dose. However, this point leads to a potential dilemma: On the one hand, a very low control dose produces little or even no effects and will therefore not prevent unblinding. On the other hand, if the control dose is high enough to maintain blinding, the efficacy of the treatment dose itself can no longer be assessed because the control dose may already be therapeutic (cf. Goodwin, Aaronson, et al. 2022). For this reason, using a lower dose of the psychedelic as an active placebo does not seem to be able to solve the blinding problem of psychedelic DB-RCTs.

The second approach to secure the logic of DB-RCTs in PAT research is to maintain blinding by anesthetizing patients during the intervention. In this way, the psychoactive effects are eliminated and patients have no indication of which group they are in. While no such study has been conducted so far with a classic serotonergic psychedelic, there is one on ketamine therapy for depression that has taken this approach (Lii et al. 2023). The authors recruited participants with major depressive disorder who were scheduled for routine surgery and either gave them a single infusion of ketamine or placebo (saline) during usual surgical anesthesia. Both ketamine and placebo infusions produced large therapeutic effects, similar to those previously published for ketamine in awake patients with depression. Importantly, blinding was successful, with only 36.8% of participants correctly guessing their treatment assignment and both groups assigning their guesses in similar proportions. The study therefore suggests that response expectancy may account for much of the antidepressant effect of ketamine. But this interpretation of the findings must still be treated with caution because neither expectations nor a possible antidepressant effect of the surgical anesthesia itself were measured.

Performing PAT in combination with anesthesia has been discussed in the PAT literature. However, the discussion has not been centered on concerns about blinding but rather on a separate debate regarding the relevance (i.e., therapeutic role or necessity) of subjectively experienced effects (Nautiyal and Yaden 2023; Yaden and Griffiths 2021). Many psychedelic researchers assume that the acute subjective experience commonly induced by psychedelics (i.e., its phenomenological aspects) are necessary for their lasting therapeutic effects. As evidence, for example, they may note that the more "mystical" the psychedelic experience as rated by patients, the greater the therapeutic effect, even when the rated intensity of the overall drug effect is statistically controlled (Yaden and Griffiths 2021). Moreover, higher reported levels of emotional breakthrough and intense responses to music-listening during a psychedelic experience are associated with greater therapeutic effects (Weiss et al. 2024). Other researchers,

however, contend that so-called non-subjective or non-hallucinogenic psychedelic analogues might be designed that could, potentially, have similar therapeutic effects to the classic (subjective) psychedelics (D. E. Olson 2021). The proposal to administer classic psychedelics in conjunction with anesthesia has then arisen in the context of this dispute. The idea is that by putting patients under anesthesia, they will have no acute subjective experience as a result of the psychedelics;[9] if they nevertheless show lasting therapeutic effects, this would imply that the subjective effects of psychedelics are not, in the end, necessary for bringing about the characteristic therapeutic outcomes. On the other hand, if anesthesia eliminates or substantially reduces therapeutic effects, as the first group of researchers anticipates, this would imply that the use of anesthesia in PAT research to blind participants would not be reasonable, as this would also eliminate a central therapeutic mechanism of psychedelics.[10]

The third approach to secure the logic of DB-RCTs in PAT research is to manipulate expectations in such a way that response expectancy does not lead to systematic bias. Specifically, therapeutic expectations must be manipulated so that they are similar for the treatment condition and the control condition. If such manipulation is successful, then unblinding will not lead to systematic bias via condition-specific response expectancy. For example, a study could compare PAT with an active control group receiving escitalopram in patients with depression. Referring to the results of Carhart-Harris et al. (2021), the experimenters could truthfully emphasize when discussing the trial with patients that both substances were equally therapeutic in a previous study. If patients adopt the belief that psilocybin and escitalopram are equally therapeutic, they will likely have similar response expectancy for the treatment group and the control group. Consequently, in the case of unblinding, there is no condition-specific response expectancy that introduces systematic bias.

Such a PAT study that tries to equalize response expectancy regarding its two conditions has not yet been conducted. However, while its results would certainly be very valuable, the design also comes with significant limitations. First, it can only tell us how therapeutic psilocybin therapy is relative to escitalopram therapy. Thus, it would not help to establish the

---

[9] Although it is not entirely clear whether this is true, because when patients are under anesthesia, psychedelics could lead to subjective effects in their dreams.

[10] This might suggest that the appropriate thing to do would be to suspend judgment on the issue until such a test is in fact conducted. And to some extent, that is a reasonable response. However, conducting psilocybin or LSD trials in combination with anesthesia is less feasible than doing so in the context of a ketamine trial because the acute subjective effects of the former drugs last longer (up to 6 hours for psilocybin and up to 12 hours for LSD), requiring longer anesthesia. The average length of surgery in the study by Lii et al. (2023) was 4 hours and 16 minutes. Prolonged exposure to anesthesia has been associated with an increased incidence of perioperative complications and mortality, especially if it exceeds 6 hours (Routh et al. 2008).

efficacy of the characteristic constituents of PAT. Second, it is unclear how effectively patients' response expectancies for the two conditions could realistically be equalized by the information that they have been equally therapeutic in a previous study. Many patients are likely influenced by the current hype surrounding psychedelics, leading them to believe that PAT is more therapeutic than other psychiatric interventions, which in turn motivates them to participate in a PAT trial (cf. Carhart-Harris et al. 2021). It is conceivable that such patients may still expect psilocybin to be more therapeutic than escitalopram, even if a study has not confirmed this. But this is ultimately an empirical question. Third, even if equalization of response expectancy were successful, systematic bias would still be possible through experimenter effects. For example, an experimenter who recognizes that a patient is in the PAT condition may (unconsciously) treat that patient more supportively than a patient they assume is in the control condition.

In fact, the possibility of such systematic bias is generally present when rescuing psychedelic DB-RCTs through patient response expectancy equalization. Only if experimenter response to unblinding is also equalized, can the logic of the DB-RCT design be truly secured. But this seems to be even more difficult to establish than patient response expectancy equalization because psychedelic researchers are already well informed about the field. In addition, they probably have some fascination with PAT, which is why they chose to do research in this area. This is not to say that for this reason psychedelic researchers are biased per se. However, there have been recent reports of potentially biased reporting of results by some psychedelic researchers, calling into question the scientific objectivity of the field (cf. Borrell 2024; Kious, Schwartz, and Lewis 2023; Mustafa et al. 2024).

To summarize the analysis of this section, it does not seem that psychedelic DB-RCTs can realistically be saved: (1) maintaining blinding through an active control group is either unsuccessful or only leads to relative results to a smaller dose of psychedelics; (2) maintaining blinding through anesthesia eliminates subjective effects which are widely, though not unanimously, thought to be necessary for the lasting therapeutic effects; and (3) equalizing expectations is difficult to achieve for both patients and experimenters, and since this method is most likely combined with a control group receiving another treatment, it does not allow for the analysis of the efficacy of the characteristic constituents of PAT.[11]

[11] Muthukumaraswamy et al. (2021) propose a fourth way to secure the logic of DB-RCTs in PAT research. They suggest using concealment to make all patients believe they are receiving a normal dose of a psychedelic under normal circumstances, but half of the patients receive a placebo, a lower dose, or pre-treatment with a substance (e.g., ketanserin) that blocks the effects of the psychedelic. Such concealment may be able to maintain patient blinding (although patients may also become suspicious in the absence of effects). However, such concealment is highly problematic from an ethical perspective, as it may invalidate informed consent by withholding information

This alone would be reason enough to conclude that it is not justified to consider the DB-RCT design as the gold standard in PAT research, as its logic is almost inevitably undermined. However, even if one disagrees with this view, and believes instead that it would merely be practically very difficult, but not impossible, to address the above-stated concerns, it could be argued that there is still another reason to reject the DB-RCT design as the gold standard in PAT research. It has to do with a term that has been used repeatedly in this paper, but not yet analyzed closely: the characteristic constituents of PAT.

5 What Makes Psychedelic-Assisted Therapy Therapeutic?

As mentioned in section 2, the main purpose of the DB-RCT design in medicine is to test whether the characteristic constituents of a treatment (as opposed to some other, incidental or confounding feature of the treatment) cause therapeutic effects. So far, we have left open what these constituents are in the case of PAT. In the literature, there are two general factions regarding this question: On the one hand, there are those who argue that it is mainly the psychedelic substance (i.e., the chemical composition of the molecules themselves) that leads to therapeutic effects (Goodwin et al. 2023). On the other hand, there are those who claim that it is the psychedelic substance *in combination with the therapeutic embedding* that leads to therapeutic effects (Alpert et al. 2024; Gründer et al. 2023; Schenberg et al. 2024). Let us examine these two positions and their implications for the DB-RCT design in PAT research more closely.

With their paper titled *Must Psilocybin Always "Assist Psychotherapy"?*, Goodwin et al. (2023) launched a controversial debate about the role of psychotherapy in PAT. The authors argue that the "therapy" provided in recent PAT studies was primarily a form of psychological support aimed at safety and not a form of psychotherapy aimed at improving efficacy. In line with that view, Goodwin et al. (2024) note that in their own study, there was no direct effect of the therapeutic alliance (i.e., the collaborative bond between therapist and patient) on the clinical outcome and only a small one on the psychedelic experience (Goodwin, Aaronson, et al. 2022; Goodwin, Simmons, et al. 2022). In addition, Goodwin et al. (2023) highlight that while the role of integration sessions following the psychedelic session has traditionally been strongly emphasized, there has been no systematic study of how much it really matters. For these

---

that is important to the patient (Eyal 2019; Villiger 2024b). For this reason, it will not be discussed further in this paper.

reasons, the authors conclude that the therapeutic effects observed to date in the best controlled PAT studies must be attributed to the psychedelic substance itself and not (also) to psychotherapy or some interaction between the two. Consequently, the term "psychedelic-assisted psychotherapy" is misleading in describing the true mechanism of change facilitated by psychedelic experience.

If Goodwin and colleagues are correct, then the standard DB-RCT design would in principle be suitable for testing the characteristic constituents of PAT (let us ignore the problems of blinding for the moment). In a standard psychedelic DB-RCT, the one difference between the treatment group and the control group is the substance administered: either it is a psychedelic or some control substance (i.e., inert placebo, active placebo, or other psychotropic drug). Everything else is held constant. Therefore, if the psychedelic itself is responsible for the therapeutic effects, any difference between the two groups can be attributed to it. The same is true if both the psychedelic itself and the psychotherapeutic embedding (or some other extra-pharmacological factors) are therapeutic, as long as these effects are simply additive. However, the same is no longer true if the therapeutic effects of PAT (partly) stem from an interaction of the psychedelic and the psychotherapeutic embedding. In this case, a standard DB-RCT cannot tell us how the psychedelic and the psychotherapeutic embedding interact with each other because it only measures the overall therapeutic effect of PAT relative to the control condition. Therefore, the standard DB-RCT design is, in principle, good at separating pharmacological effects from extra-pharmacological effects, but only when they do not interact. According to Goodwin and colleagues, there is no such interaction in PAT, which makes the standard DB-RCT—again, in principle, setting aside issues around blinding, and so on—a suitable design to study the therapeutic mechanisms of PAT.

However, many authors disagree with the position of Goodwin and colleagues (Alpert et al. 2024; Deckel, Lepow, and Guss 2024; Dworkin et al. 2023; Earleywine et al. 2024; Gründer et al. 2023; Schenberg et al. 2024). They highlight that there is no clear evidence that the psychedelic substance alone causes the therapeutic effects of PAT; current data do not allow us to neatly dissect the effects of the psychedelic substance, the therapy, and their interaction. At the same time, there is ample evidence that the set and setting of a psychedelic trip influence its manifestations and its effects, demonstrating the importance of extra-pharmacological factors (Hartogsohn 2016; 2020). In fact, some authors even argue that psychedelics are "active super-placebos," in the sense that they are mainly catalysts and amplifiers of what is already there (e.g., pre-existing beliefs, intentions, motivations, and so on) (Dupuis and Veissière 2022; Grof 2008; Hartogsohn 2020; Langlitz et al. 2021; Pollan 2018; Strassman 2017; Weil 1972). Building on that supposition, Villiger (2022) developed an account of the interaction between

psychedelics and psychotherapy which holds that psychedelics enhance the therapeutic effects of psychotherapy by amplifying its common factors. These common factors include the *real relationship*, i.e., "the personal relationship between therapist and patient marked by the extent to which each is genuine with the other and perceives/experiences the other in ways that befit the other" (Gelso 2014, 119); *expectations*, i.e., the patient's expectations regarding the treatment; and *specific ingredients*, i.e., the therapeutic actions of the treatment (cf. Wampold and Imel 2015).

Whether psychedelics *do* amplify the common factors of psychotherapy is still an open question, but there are indications for such an interaction. First, contrary to the findings of Goodwin et al. (2022), a recent analysis of the role of therapeutic alliance in PAT emphasizes its relevance (Levin et al. 2024): Therapeutic alliance at the final preparation session and one week after the psychedelic session correlated moderately to strongly with depression scores at 4 weeks, 6 months, and 12 months post-intervention. Furthermore, therapeutic alliance increased significantly from the final preparation session to one week after the psychedelic session. Finally, therapeutic alliance at the final preparation session was moderately correlated with ratings of mystical experience and psychological insight (in turn, mystical experience and psychological insight were moderately and strongly correlated with depression scores at 4 weeks post-intervention, respectively). These findings suggest that therapeutic alliance contributes to the therapeutic effects of PAT and that there is likely an interaction between therapeutic alliance and the psychedelic experience.

Importantly, therapeutic alliance is directly related to the common factors of psychotherapy as it is about the bond between the patient and the therapist (connected to the real relationship), about the goals of treatment (connected to expectations), and the tasks of therapy (connected to specific ingredients). Therefore, the study by Levin et al. suggests that psychedelics boost common factors and thereby therapeutic effects via improving therapeutic alliance. Consistent with this, a qualitative study of PAT found that the psychedelic session strengthened the patient's bond with and trust in the therapist, which is likely to enhance the real relationship (Watts et al. 2017). Still, it should also be noted that the studies on response expectancy in PAT discussed in Section 3 provide mixed evidence for a positive effect of therapeutic expectations on treatment outcomes (Szigeti et al. 2024; Weiss et al. 2021). Thus, the evidence that expectations as a common factor are important in PAT and interact with the psychedelic substance remains inconclusive. Nevertheless, the current body of evidence does not allow the conclusion that there is *no* interaction between the psychedelic substance and the psychotherapeutic embedding in PAT.

Taken together, it remains possible, if not likely, that the characteristic constituents of PAT should be a combination of the psychedelic substance, the psychotherapeutic embedding, and their interaction. This leads to the following problem: in the context of PAT, the standard DB-RCT with a psychedelic condition and a control condition cannot shed light on this issue because it is unable to distinguish between these effects.[12] Even worse, because of its usual focus on pharmacological effects, the standard DB-RCT design reduces attention to extra-pharmacological effects and interactions between effects and, consequently, how to optimize them. For example, although PAT consists of both a pharmacological and a psychological intervention, there is very little empirical research on the psychological part of PAT and/or on the interaction between the two interventions (cf. McNamee, Devenot, and Buisson 2023). Put differently, PAT has mainly been studied as a pharmacological intervention so far, and this seems to be related to the fact that DB-RCTs are still considered the gold standard in PAT research. Only by rejecting the DB-RCT design as the gold standard, and instead considering DB-RCTs as complementary—rather than superior—to trials using other research designs, can we overcome the limited perspective of the standard DB-RCT and study PAT more holistically. So, in a final step, let's analyze for which questions the DB-RCT design is still useful, which research designs complement it in answering other questions, and how combining the results from these different research designs allows us to get a more complete picture of PAT.

6 How Psychedelic-Assisted Therapy Should Be Studied

We begin by analyzing where the DB-RCT design remains useful in PAT research, starting with DB-RCTs that manipulate the substance administered. First, DB-RCTs that use a lower psychedelic dose as the control condition are helpful in determining whether a higher dose increases therapeutic effects to a clinically significant degree. This is especially true if the dose in the control condition is still high enough to produce psychoactive effects sufficient to maintain blinding. For example, in an LSD study that used different doses (25 µg, 50 µg, 100 µg, and 200 µg), participants were able to identify the 100 and 200 µg doses as high doses but unable to distinguish between them (Holze et al. 2021). Thus, in a DB-RCT using these two

[12] Ignoring the problems of blinding, one would need a DB-RCT with four conditions to distinguish between these effects: (1) psychedelic with psychotherapy; (2) psychedelic without psychotherapy; (3) placebo with psychotherapy; and (4) placebo without psychotherapy.

doses, differences in therapeutic outcomes between the two groups should be attributable to the difference in dosing.[13]

Second, DB-RCTs that use another treatment as the control condition are helpful in determining whether PAT performs better than this other treatment. This is particularly true if response expectancy is equalized across the two conditions because doing so minimizes systematic bias introduced by patient unblinding (however, as discussed in Section 4, response expectancy equalization is difficult to achieve). Then again, one might ask, what is the point of blinding here if we acknowledge that unblinding is, for all intents and purposes, inevitable after the substances are administered? First, blinding balances out anticipatory effects that are specific to one of the two treatments. Second, blinding has the advantage that patients who enter the study because they want to undergo PAT will not drop out before the "psychedelic session" and are likely to complete the study afterwards (cf. Carhart-Harris et al. 2021; Goodwin, Aaronson, et al. 2022; Rotz et al. 2023). In contrast, if these same patients learn at the outset that they are in the control group, they may have little motivation to continue in the study and drop out early, leading to (stronger) sampling bias.

Third, placebo-controlled DB-RCTs where the psychedelic session is performed under anesthesia might be helpful in determining how important subjective effects and response expectancy are in PAT.[14] On the one hand, if it turned out that there was a meaningful therapeutic difference between PAT and placebo under anesthesia (favoring PAT), this would imply that PAT can be therapeutic without subjective effects. Moreover, if the size of the therapeutic effects of PAT in such a study turned out to be similar to those in other PAT studies, this would be strong evidence that subjective effects are indeed irrelevant for lasting therapeutic effects. Alternatively, if smaller than usual effect sizes would be observed, this could suggest at least *some* therapeutic relevance of subjective effects. However, it might also be possible that smaller effect sizes are due to lower response expectancy, meaning that patients expect PAT to be less therapeutic under anesthesia. It would therefore be necessary to assess potential differences in response expectancy here. On the other hand, if it turned out that there is no therapeutic difference between PAT under anesthesia and placebo under anesthesia, this would have two possible implications: If effect sizes are similar to those in other PAT studies, subjective effects seem irrelevant for lasting therapeutic effects and response expectancy seems crucial in PAT. Conversely, if effect sizes are smaller than those in other PAT studies, either the pharmacological

[13] Note, however, that assuming the study finds such a difference, it remains unclear whether it is due to pharmacological factors, the interaction between pharmacological and extra-pharmacological factors, or both.

[14] Such a study should be conducted with DMT or psilocybin as they would not require anesthesia for more than 6 hours.

intervention without subjective effects does not appear to be therapeutic, or subjective effects appear to be irrelevant and PAT under anesthesia produces lower response expectancy. Thus, potential differences in response expectancy need to be assessed here as well. Finally, in all these effect size comparisons between PAT under anesthesia and PAT under normal conditions, it is important to exclude that anesthesia itself has therapeutic effects.

While standard psychedelic DB-RCTs manipulate the substance administered, other manipulations are also possible. Studies including such manipulations should help us better understand the role of extra-pharmacological factors in PAT. First, psychedelic DB-RCTs could manipulate *expectations*. That is, both groups undergo PAT, but they receive different information beforehand, in the hope that this will induce different therapeutic expectations. For example, one group could receive more critical information about PAT than the other (e.g., that there is currently a kind of hype around psychedelics and that therapeutic effects have likely been overestimated in past studies). Differences in treatment outcomes between the two groups could then be attributed to the different information they received and would reveal the presence of expectancy effects. To maintain blinding, or to do so as much as possible, patients could receive vague information about the study design (e.g., that the study examines whether extra-pharmacological factors influence treatment outcomes without mentioning that there is a manipulation).[15]

Second, psychedelic DB-RCTs could manipulate the *therapeutic embedding*. For instance, the two conditions could be psychological support vs. psychotherapy;[16] psychological support vs. psychological support with a focus on the therapeutic alliance; standard preparatory sessions vs. extended preparatory sessions; psychedelic session in a familiar context (i.e., same experimenter/room as in preparatory sessions) vs. psychedelic session in an unfamiliar context; psychedelic session in an enriched context (e.g., music available, soft lightning, comfortable room) vs. psychedelic session in an unenriched context (cf. Carhart-Harris et al. 2018); and so on. Again, to maintain participant blinding, patients should be given vague information about the study design so that the fact of manipulation is not revealed. At the same time, it will usually not be possible to maintain experimenter blinding (it may be possible if experimenters are unaware of the manipulation, treat only patients of the same condition, and treat only a few of the

[15] While experimenter blinding should also be attempted here, unblinding might still occur if a patient specifically referred to the manipulated information in conversation with the experimenter.

[16] Of course, there is not just one type of psychotherapy, but many. So, the type of psychotherapy (e.g., cognitive-behavioral or client-centered) would have to be specified. Consequently, the two conditions of a psychedelic DB-RCT could also be Psychotherapy A vs. Psychotherapy B.

total number of patients). Therefore, systematic bias introduced by experimenters cannot be ruled out.

Let us continue with designs beyond the DB-RCT. One way to measure the therapeutic effects of PAT is to have an RCT with a treatment group and a wait-list control group. The wait-list control group controls for Hawthorne effects, spontaneous remission, and regression to the mean. Thus, the difference between the treatment group and the control group should largely capture the therapeutic effects of PAT so long as we assume that extra-pharmacological factors are also characteristic constituents of PAT. Such a design has three major advantages compared to a DB-RCT design. First, it has higher external validity because patients outside the research context do not have to worry about whether or not they will receive the treatment they want. Second, patients in the control group have a strong incentive to stay in the study because if they do, they will eventually receive PAT. Third, the design is more efficient because the control group does not receive any placebo treatment (i.e., preparation session, sham psychedelic session, and integration session).

Despite these advantages, such an open-label design also comes with problems: patients in the wait-list control group may be disappointed about the wait, which may have a negative impact on their condition and thereby increase the difference between the two groups (but anticipatory effects caused by the fact that they will receive PAT eventually may also counteract such a disappointment effect). Furthermore, patients in a wait-list control group may wait a long time until they get some treatment, which can be problematic from an ethical perspective. Admittedly, this is also the case for placebo-controlled studies. In fact, it is even worse there, because typically patients do not even have the opportunity to receive PAT after the trial is completed. However, in contrast to wait-list or placebo-controlled trials, all patients in trials that use another treatment as the control condition receive some treatment, so no one is left untreated for the duration of the trial. This seems to be ethically favorable. Then again, it is also possible that patients would prefer to receive PAT with certainty either during or after the trial rather than either PAT or another treatment during the trial at 50% each. So, in the end, it is not entirely clear which of the two options is more ethical—at least not when the only chance of receiving PAT is to participate in a trial and patients that receive another treatment as part of the trial cannot choose to receive PAT after the trial is completed.

The pragmatic clinical trial (PCT) design provides another method to study the efficacy of PAT, but with a different purpose than the RCT design. PCTs focus on external validity, meaning that they examine whether PAT is therapeutic not only in tightly controlled settings, but also under more real-world conditions. Accordingly, unlike RCTs, PCTs are not explanatory/confirmatory trials that analyze the therapeutic mechanisms of a treatment. That is why

PCTs are usually conducted after the efficacy of a treatment has already been shown in RCTs and the treatment has reached phase 4, also called the post-marketing phase. But Carhart-Harris et al. (2022) have argued for the potential benefits of using PCTs earlier to comprehensively assess the effectiveness of PAT.

According to Carhart-Harris et al., a major advantage of PCTs is that they allow more flexibility in the study protocol, such as flexibility in dosage, number of psychedelic sessions, and number of integration sessions. This helps to determine which conditions are optimal and whether there may be individual differences in this regard. Relatedly, because PAT studies to date have had relatively small and homogeneous samples, it is unclear whether there are patient characteristics that predict a better response to PAT. In fact, it is even unclear whether the results of previous PAT studies can be extrapolated to the heterogeneous population of potential PAT patients due to the highly selected samples. Finally, Carhart-Harris et al. (2022) propose that psychedelic PCTs be combined with so-called basket protocols. Basket protocols are single protocols that have been approved by relevant research regulatory bodies for the testing of a single intervention for several conditions. In this way, the same PAT intervention could be tested for multiple mental health conditions, which is efficient from a research perspective and counteracts the over-fragmentation of PAT.

PCTs can also include a control group, such as a group of patients who chose a treatment other than PAT. This allows the efficacy of the two treatments to be compared under real-world conditions. A beneficial aspect of such a design is that patients choose the treatment they receive, which eliminates possible disappointment effects associated with randomization. In addition, no one has to wait until the end of the trial to receive the treatment, which is an advantage over a wait-list controlled trial.

The strength of PCTs, namely external validity, also poses a potential danger for participants. With less stringent sample selection based on safety criteria, studies may result in more serious adverse events (for criteria of a serious adverse event, see FDA 2023). We already find such indications in the three existing phase 2 trials: Both Carhart-Harris et al. (2021) and Raison et al. (2023) had highly restrictive selection criteria: only 6-7% of the screened participants made it into the final study. There were no serious adverse events in either study. In contrast, the sample of Goodwin et al. (2022) seems to be more inclusive as 54% of the screened participants made it into the final study. Here, approximately 5% of participants in the treatment groups (10 mg or 25 mg of psilocybin) experienced a serious adverse event after the psychedelic session. While the reason for this higher number of serious adverse events is not entirely clear, Goodwin et al. (2024) suggest that it is due to a less selective sample. This seems plausible and would imply that an increase in external validity through less restrictive inclusion criteria could

be associated with a greater overall risk to participants. PCTs need to consider this trade-off. Still, if PAT is therapeutic with highly selected samples, as preliminary evidence currently suggests, sooner or later the inclusion criteria will have to be (slowly) relaxed to see if the results can be replicated with more representative samples.

As with all the study designs discussed in this section, PAT studies should always assess the following variables, preferably before and after the psychedelic session: (1) therapeutic expectations about PAT and about the control treatment if there is one (e.g., through the Treatment Expectation Questionnaire; see Alberts et al. 2020); (2) therapeutic alliance (e.g., through the Working Alliance Inventory-Short Revised; see Hatcher and Gillaspy 2006); and (3) therapist empathy, which is an indicator for the real relationship (e.g., through the empathy scale of the Barrett-Lennard Relationship Inventory: Other to Self version; see Barrett-Lennard 2015). The evaluation of these variables makes it possible to see if the therapeutic outcomes of PAT are correlated with common factors of psychotherapy and if the psychedelic session interacts with them (cf. Villiger 2022). Such analysis is crucial for the debate on the characteristic constituents of PAT. Finally, if there has been blinding, there should always be an evaluation of how successful it has been (e.g., through Bang's blinding index; see Bang, Ni, and Davis 2004).

The combined results of studies using the different research designs presented above provide a more complete picture of PAT. In this picture, DB-RCTs play an important role because they allow causal inferences to be made about the effects of different dosages, therapeutic expectations, and the therapeutic embedding; they allow researchers to analyze the relevance of subjective effects and response expectancy; and they allow a comparison between PAT and some other treatment that controls for as many confounding variables as possible (but not all confounding variable, making causal inference difficult).

However, when evaluating the scientific validity of PAT, the results of DB-RCTs should not be seen as superior but complementary to the results of trials using other research designs. RCTs with a wait-list control group allow to evaluate the efficacy of PAT when both pharmacological and extra-pharmacological aspects are considered constitutive of PAT. And PCTs allow researchers to evaluate (and compare) PAT under more real-world conditions, to analyze the optimal (personal) PAT procedure, and to assess whether the same study protocol works for different mental disorders. Finally, the inclusion of variables measuring common factors allows for the investigation of the extent to which common factors correlate with therapeutic effects and whether there is an interaction between common factors and the psychedelic experience.

Taken together, we see there is a kind of labor division between the different research designs (see Table 1). An RCT with a treatment group and a wait-list control group is the best design to test whether PAT is therapeutic from an absolute perspective, because only here is

the control group not exposed to the extra-pharmacological factors of PAT (or other therapeutic factors). These findings can then be complemented with those from a relative perspective, namely a DB-RCT with another treatment as the control group. But it should be noted that in such a trial, the logic of the DB-RCT will be undermined due to unsuccessful blinding, so causal inferences are limited. In contrast, to answer the question of what makes PAT therapeutic, we need causal inferences and thus DB-RCTs with successful blinding. Each factor that could make PAT therapeutic must be tested individually, on the condition that successful blinding is possible. Finally, to answer whether PAT is therapeutic not only in tightly controlled settings but also under more real-world conditions, and to find the optimal configurations of PAT, we need PCTs.

Table 1: How to gain a more complete picture of PAT.

Is PAT therapeutic (under tightly controlled conditions)?	**What makes PAT therapeutic?**	**Is PAT therapeutic under real-world conditions?**
RCT with a wait-list control group can tell us whether PAT is therapeutic, assuming that both pharmacological and extra-pharmacological aspects are constitutive of PAT.	DB-RCT with different doses of the psychedelic can tell us whether a higher dose increases the therapeutic effect.	PCTs with more heterogeneous samples can tell us whether PAT is (also) therapeutic in samples with less restrictive inclusion criteria.
(DB-)RCT with another treatment as the control condition can tell us how PAT compares to that other treatment (note that maintaining blinding will be unsuccessful here).	DB-RCTs manipulating extra-pharmacological factors (e.g., expectations, therapeutic embedding) can tell us to what extent extra-pharmacological factors influence therapeutic effects.	PCTs with more flexible study protocols can tell us which configurations are optimal and whether there are individual differences in this regard.
	DB-RCT under anesthesia can tell us about the relevance of subjective experience during a psychedelic session and the relevance of response expectancy.	PCTs with basket protocols for several conditions can tell us whether the same type of PAT intervention can be used for multiple mental health conditions.
		PCTs with a control group getting another treatment can tell us how PAT compares to that other treatment under real-world conditions.

Note. All studies should include measurements of: (1) therapeutic expectations; (2) therapeutic alliance; (3) therapist empathy; (4) blinding success (if applicable).

Ideally, the standard DB-RCT with a treatment condition and an (active) placebo condition answers both whether the treatment is therapeutic and what makes it therapeutic—this is the inherent promise of the DB-RCT gold standard. Unfortunately, as shown above, such a standard DB-RCT is of limited use in the context of PAT as it cannot properly answer either question. For this, we need several studies with different research designs. On the one hand, studies that aim to answer what makes PAT therapeutic must follow the DB-RCT design and maintain blinding, which precludes the standard DB-RCT. On the other hand, studies that aim to answer whether PAT is therapeutic cannot maintain blinding (unless they use a control group with a lower but still significant psychedelic dose, leading to relative results only). Accordingly, the strength of the DB-RCT design—its internal validity—is undermined when we use it to test whether PAT is therapeutic. This justifies why RCTs with a wait-list control group and PCTs, which both focus less on internal and more on external validity, are not inherently inferior in testing whether PAT is therapeutic compared to DB-RCTs.

7 Conclusion

This paper examined whether the DB-RCT design, commonly considered the gold standard in medical research, should also be the gold standard in PAT research. It presented the logic behind the DB-RCT design and showed that this logic is in danger of being undermined in PAT research. Further analysis revealed that salvaging the DB-RCT design to enable it to determine whether PAT is therapeutic is not feasible. Nor is it desirable because standard DB-RCTs neglect the extra-pharmacological factors of PAT and their potential interactions with the pharmacological factors; both of which are considered by many to be constitutive of PAT. For these reasons, we should abandon the idea that the DB-RCT design should be considered the gold standard in PAT research. This allows us to study PAT more holistically, where DB-RCTs are still useful, especially to examine specific aspects of PAT, but their results are not superior but complementary to those of randomized wait-list controlled trials or PCTs.

References

Aday, Jacob S., Robin L. Carhart-Harris, and Joshua D. Woolley. 2023. “Emerging Challenges for Psychedelic Therapy.” *JAMA Psychiatry*, April. https://doi.org/10.1001/jamapsychiatry.2023.0549.

Aday, Jacob S., Boris D. Heifets, Steven D. Pratscher, Ellen Bradley, Raymond Rosen, and Joshua D. Woolley. 2022. "Great Expectations: Recommendations for Improving the Methodological Rigor of Psychedelic Clinical Trials." *Psychopharmacology* 239 (6): 1989–2010. https://doi.org/10.1007/s00213-022-06123-7.

Aday, Jacob S., Julia R. Wood, Emily K. Bloesch, and Christopher C. Davoli. 2021. "Psychedelic Drugs and Perception: A Narrative Review of the First Era of Research." *Reviews in the Neurosciences* 32 (5): 559–71. https://doi.org/10.1515/revneuro-2020-0094.

Alberts, Jannis, Bernd Löwe, Maja Alicia Glahn, Keith Petrie, Johannes Laferton, Yvonne Nestoriuc, and Meike Shedden-Mora. 2020. "Development of the Generic, Multidimensional Treatment Expectation Questionnaire (TEX-Q) through Systematic Literature Review, Expert Surveys and Qualitative Interviews." *BMJ Open* 10 (8): e036169. https://doi.org/10.1136/bmjopen-2019-036169.

Alpert, Michael D., Kelley C. O'Donnell, Casey A. Paleos, Evan Sola, Christopher S. Stauffer, Anne C. Wagner, Christopher R. Nicholas, and Michael C. Mithoefer. 2024. "Psychotherapy in Psychedelic Treatment: Safe, Evidence-Based, and Necessary." *American Journal of Psychiatry* 181 (1): 76–77. https://doi.org/10.1176/appi.ajp.20230665.

Bang, Heejung, Liyun Ni, and Clarence E. Davis. 2004. "Assessment of Blinding in Clinical Trials." *Controlled Clinical Trials* 25 (2): 143–56.

Barrett-Lennard, G. T. 2015. *The Relationship Inventory: A Complete Resource and Guide*. Hoboken, NJ: John Wiley & Sons.

Bender, David, and David J. Hellerstein. 2022. "Assessing the Risk–Benefit Profile of Classical Psychedelics: A Clinical Review of Second-Wave Psychedelic Research." *Psychopharmacology* 239 (6): 1907–32. https://doi.org/10.1007/s00213-021-06049-6.

Blease, Charlotte, Michael H Bernstein, and Cosima Locher. 2020. "Open-Label Placebo Clinical Trials: Is It the Rationale, the Interaction or the Pill?" *Primary Care* 25 (5): 7.

Bogenschutz, Michael P., Alyssa A. Forcehimes, Jessica A. Pommy, Claire E. Wilcox, P. C. R. Barbosa, and Rick J. Strassman. 2015. "Psilocybin-Assisted Treatment for Alcohol Dependence: A Proof-of-Concept Study." *Journal of Psychopharmacology (Oxford, England)* 29 (3): 289–99. https://doi.org/10.1177/0269881114565144.

Borrell, Brendan. 2024. "The Psychedelic Evangelist." *The New York Times*, March 21, 2024, sec. Health. https://www.nytimes.com/2024/03/21/health/psychedelics-roland-griffiths-johns-hopkins.html.

Butler, Matt, Luke Jelen, and James Rucker. 2022. "Expectancy in Placebo-Controlled Trials of Psychedelics: If so, so What?" *Psychopharmacology* 239 (10): 3047–55. https://doi.org/10.1007/s00213-022-06221-6.

Carhart-Harris, Robin L., Bruna Giribaldi, Rosalind Watts, Michelle Baker-Jones, Ashleigh Murphy-Beiner, Roberta Murphy, Jonny Martell, Allan Blemings, David Erritzoe, and David J. Nutt. 2021. "Trial of Psilocybin versus Escitalopram for Depression." *New England Journal of Medicine* 384 (15): 1402–11. https://doi.org/10.1056/NEJMoa2032994.

Carhart-Harris, Robin L., Mendel Kaelen, Mark Bolstridge, T. M. Williams, L. T. Williams, Raphael Underwood, Amanda Feilding, and David J. Nutt. 2016. "The Paradoxical Psychological Effects of Lysergic Acid Diethylamide (LSD)." *Psychological Medicine* 46 (7): 1379–90.

Carhart-Harris, Robin L., Leor Roseman, Eline Haijen, David Erritzoe, Rosalind Watts, Igor Branchi, and Mendel Kaelen. 2018. "Psychedelics and the Essential Importance of Context." *Journal of Psychopharmacology* 32 (7): 725–31.

Carhart-Harris, Robin L, Anne C Wagner, Manish Agrawal, Hannes Kettner, Jerold F Rosenbaum, Adam Gazzaley, David J Nutt, and David Erritzoe. 2022. "Can Pragmatic Research, Real-World Data and Digital Technologies Aid the Development of

Psychedelic Medicine?" *Journal of Psychopharmacology* 36 (1): 6–11. https://doi.org/10.1177/02698811211008567.
ClinicalTrials. 2024. "Ongoing Clinical Trials with Psilocybin." 2024. https://clinicaltrials.gov./search?intr=psilocybin&aggFilters=status:not%20rec%20act.
Davidson, Richard J., and Alfred W. Kaszniak. 2015. "Conceptual and Methodological Issues in Research on Mindfulness and Meditation." *American Psychologist* 70 (7): 581–92. https://doi.org/10.1037/a0039512.
Deckel, Garrett Marie, Lauren A. Lepow, and Jeffrey Guss. 2024. "'Psychedelic Assisted Therapy' Must Not Be Retired." *American Journal of Psychiatry* 181 (1): 77–78. https://doi.org/10.1176/appi.ajp.20230667.
Dos Santos, Rafael G., Flávia de Lima Osório, Juliana Mendes Rocha, Giordano Novak Rossi, José Carlos Bouso, Lucas S. Rodrigues, Gabriela de Oliveira Silveira, Mauricio Yonamine, and Jaime E. Cecílio Hallak. 2021. "Ayahuasca Improves Self-Perception of Speech Performance in Subjects with Social Anxiety Disorder: A Pilot, Proof-of-Concept, Randomized, Placebo-Controlled Trial." *Journal of Clinical Psychopharmacology* 41 (5): 540–50.
Dupuis, David, and Samuel Veissière. 2022. "Culture, Context, and Ethics in the Therapeutic Use of Hallucinogens: Psychedelics as Active Super-Placebos?" *Transcultural Psychiatry* 59 (5): 571–78. https://doi.org/10.1177/13634615221131465.
Dworkin, Robert H., Michael P. McDermott, Sandeep M. Nayak, and Eric C. Strain. 2023. "Psychedelics and Psychotherapy: Is the Whole Greater than the Sum of Its Parts?" *Clinical Pharmacology & Therapeutics* 114 (6): 1166–69. https://doi.org/10.1002/cpt.3050.
Dyck, Erika. 2008. *Psychedelic Psychiatry: LSD from Clinic to Campus*. Baltimore, Md: Johns Hopkins University Press.
Earleywine, Mitch, Joseph De Leo, Dinesh Bhayana, Bhavya Rajanna, and Karen Scott. 2024. "Psilocybin Without Psychotherapy: A Cart Without a Horse?" *American Journal of Psychiatry* 181 (1): 78–79. https://doi.org/10.1176/appi.ajp.20230572.
Eyal, Nir. 2019. "Informed Consent." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2019. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2019/entries/informed-consent/.
FDA. 2023. "What Is a Serious Adverse Event?" What Is a Serious Adverse Event? FDA. 2023. https://www.fda.gov/safety/reporting-serious-problems-fda/what-serious-adverse-event.
Gaab, Jens, Cosima Locher, and Manuel Trachsel. 2021. "The Ethics of Placebo and Nocebo in Psychotherapy." In *The Oxford Handbook of Psychotherapy Ethics*, edited by Manuel Trachsel, Jens Gaab, Nikola Biller-Andorno, Şerife Tekin, and John Z. Sadler, First edition, 398–411. Oxford Handbooks in Philosophy and Psychiatry. New York, NY: Oxford University Press.
Gasser, Peter, Dominique Holstein, Yvonne Michel, Rick Doblin, Berra Yazar-Klosinski, Torsten Passie, and Rudolf Brenneisen. 2014. "Safety and Efficacy of Lysergic Acid Diethylamide-Assisted Psychotherapy for Anxiety Associated with Life-Threatening Diseases." *The Journal of Nervous and Mental Disease* 202 (7): 513.
Gelso, Charles J. 2014. "A Tripartite Model of the Therapeutic Relationship: Theory, Research, and Practice." *Psychotherapy Research* 24 (2): 117–31.
Goodwin, Guy M., Scott T. Aaronson, Oscar Alvarez, Peter C. Arden, Annie Baker, James C. Bennett, Catherine Bird, et al. 2022. "Single-Dose Psilocybin for a Treatment-Resistant Episode of Major Depression." *New England Journal of Medicine* 387 (18): 1637–48. https://doi.org/10.1056/NEJMoa2206443.
Goodwin, Guy M., Ekaterina Malievskaia, Gregory A. Fonzo, and Charles B. Nemeroff. 2023. "Must Psilocybin Always 'Assist Psychotherapy'?" *American Journal of Psychiatry* 181 (1): 20–25. https://doi.org/10.1176/appi.ajp.20221043.

———. 2024. “Psychological Support for Psilocybin Treatment: Reply to Letters on Our Commentary.” *American Journal of Psychiatry* 181 (1): 79–81. https://doi.org/10.1176/appi.ajp.20230905.
Goodwin, Guy M., Hollie Simmons, Lindsey Marwood, Sunil Mistry, Danielle Schlosser, Joyce Tsai, and Ekaterina Malievskaia. 2022. “Of Therapeutic Alliance and the Psychedelic Experience Using Path Modeling in a Phase IIb Randomized Controlled Trial of COMP360 Psilocybin Therapy.” In . Phoenix.
Griffiths, Roland R., William A. Richards, Una McCann, and Robert Jesse. 2006. “Psilocybin Can Occasion Mystical-Type Experiences Having Substantial and Sustained Personal Meaning and Spiritual Significance.” *Psychopharmacology* 187 (3): 268–83.
Grinspoon, L., and J. B. Bakalar. 1997. *Marihuana, the Forbidden Medicine*. New Haven: Yale University Press.
Grob, Charles S., Alicia L. Danforth, Gurpreet S. Chopra, Marycie Hagerty, Charles R. McKay, Adam L. Halberstadt, and George R. Greer. 2011. “Pilot Study of Psilocybin Treatment for Anxiety in Patients With Advanced-Stage Cancer.” *Archives of General Psychiatry* 68 (1): 71–78. https://doi.org/10.1001/archgenpsychiatry.2010.116.
Grof, Stanislav. 2008. *LSD Psychotherapy*. 4th ed. San Jose, CA: Multidisciplinary Association for Psychedelic Studies.
Grünbaum, Adolf. 1981. “The Placebo Concept.” *Behaviour Research and Therapy* 19 (2): 157–67. https://doi.org/10.1016/0005-7967(81)90040-1.
Gründer, Gerhard, Manuela Brand, Lea J. Mertens, Henrik Jungaberle, Laura Kärtner, Dennis J. Scharf, Moritz Spangemacher, and Max Wolff. 2023. “Treatment with Psychedelics Is Psychotherapy: Beyond Reductionism.” *The Lancet Psychiatry* 11 (3): 231–36. https://doi.org/10.1016/S2215-0366(23)00363-2.
Gupta, Snehil, and Anju Dhawan. 2022. “Methodological Issues in Conducting Yoga- and Meditation-Based Research: A Narrative Review and Research Implications.” *Journal of Ayurveda and Integrative Medicine* 13 (3): 100620. https://doi.org/10.1016/j.jaim.2022.100620.
Hartogsohn, Ido. 2016. “Set and Setting, Psychedelics and the Placebo Response: An Extra-Pharmacological Perspective on Psychopharmacology.” *Journal of Psychopharmacology* 30 (12): 1259–67.
———. 2020. *American Trip: Set, Setting, and the Psychedelic Experience in the Twentieth Century*. Cambridge, Massachusetts: The MIT Press.
Hatcher, Robert L., and J. Arthur Gillaspy. 2006. “Development and Validation of a Revised Short Version of the Working Alliance Inventory.” *Psychotherapy Research* 16 (1): 12–25. https://doi.org/10.1080/10503300500352500.
Holze, Friederike, Peter Gasser, Felix Müller, Patrick C. Dolder, and Matthias E. Liechti. 2023. “Lysergic Acid Diethylamide–Assisted Therapy in Patients with Anxiety with and without a Life-Threatening Illness: A Randomized, Double-Blind, Placebo-Controlled Phase II Study.” *Biological Psychiatry* 93 (3): 215–23.
Holze, Friederike, Patrick Vizeli, Laura Ley, Felix Müller, Patrick Dolder, Melanie Stocker, Urs Duthaler, et al. 2021. “Acute Dose-Dependent Effects of Lysergic Acid Diethylamide in a Double-Blind Placebo-Controlled Study in Healthy Subjects.” *Neuropsychopharmacology* 46 (3): 537–44. https://doi.org/10.1038/s41386-020-00883-6.
Hovmand, Oliver Rumle, Emil Deleuran Poulsen, Sidse Arnfred, and Ole Jakob Storebø. 2023. “Risk of Bias in Randomized Clinical Trials on Psychedelic Medicine: A Systematic Review.” *Journal of Psychopharmacology* 37 (7): 649–59. https://doi.org/10.1177/02698811231180276.
How to Change Your Mind. 2022. Documentary. Jigsaw Productions, Tree Tree Tree.
Howick, Jeremy. 2011. *The Philosophy of Evidence-Based Medicine*. Chichester, UK: John Wiley & Sons.

Johnson, Matthew W., Albert Garcia-Romeu, and Roland R. Griffiths. 2017. “Long-Term Follow-up of Psilocybin-Facilitated Smoking Cessation.” *The American Journal of Drug and Alcohol Abuse* 43 (1): 55–60.

Johnson, Matthew W., Roland R. Griffiths, Peter S. Hendricks, and Jack E. Henningfield. 2018. “The Abuse Potential of Medical Psilocybin According to the 8 Factors of the Controlled Substances Act.” *Neuropharmacology*, Psychedelics: New Doors, Altered Perceptions, 142 (November):143–66. https://doi.org/10.1016/j.neuropharm.2018.05.012.

Kaptchuk, Ted J. 1998. “Intentional Ignorance: A History of Blind Assessment and Placebo Controls in Medicine.” *Bulletin of the History of Medicine* 72 (3): 389–433.

Kaptchuk, Ted J. 2001. “The Double-Blind, Randomized, Placebo-Controlled Trial: Gold Standard or Golden Calf?” *Journal of Clinical Epidemiology*.

Kious, Brent, Zach Schwartz, and Benjamin Lewis. 2023. “Should We Be Leery of Being Leary? Concerns about Psychedelic Use by Psychedelic Researchers.” *Journal of Psychopharmacology* 37 (1): 45–48. https://doi.org/10.1177/02698811221133461.

Langlitz, Nicolas, Erika Dyck, Milan Scheidegger, and Dimitris Repantis. 2021. “Moral Psychopharmacology Needs Moral Inquiry: The Case of Psychedelics.” *Frontiers in Psychiatry* 12. https://www.frontiersin.org/articles/10.3389/fpsyt.2021.680064.

Levin, Adam W., Rafaelle Lancelotta, Nathan D. Sepeda, Natalie Gukasyan, Sandeep Nayak, Theodore L. Wagener, Frederick S. Barrett, Roland R. Griffiths, and Alan K. Davis. 2024. “The Therapeutic Alliance between Study Participants and Intervention Facilitators Is Associated with Acute Effects and Clinical Outcomes in a Psilocybin-Assisted Therapy Trial for Major Depressive Disorder.” *PloS One* 19 (3): e0300501.

Lii, Theresa R., Ashleigh E. Smith, Josephine R. Flohr, Robin L. Okada, Cynthia A. Nyongesa, Lisa J. Cianfichi, Laura M. Hack, Alan F. Schatzberg, and Boris D. Heifets. 2023. “Randomized Trial of Ketamine Masked by Surgical Anesthesia in Patients with Depression.” *Nature Mental Health* 1 (11): 876–86. https://doi.org/10.1038/s44220-023-00140-x.

Locher, Cosima, Jens Gaab, and Charlotte Blease. 2018. “When a Placebo Is Not a Placebo: Problems and Solutions to the Gold Standard in Psychotherapy Research.” *Frontiers in Psychology* 9. https://doi.org/10.3389/fpsyg.2018.02317.

Loo, Colleen, Nick Glozier, David Barton, Bernhard T. Baune, Natalie T. Mills, Paul Fitzgerald, Paul Glue, et al. 2023. “Efficacy and Safety of a 4-Week Course of Repeated Subcutaneous Ketamine Injections for Treatment-Resistant Depression (KADS Study): Randomised Double-Blind Active-Controlled Trial.” *The British Journal of Psychiatry* 223 (6): 533–41. https://doi.org/10.1192/bjp.2023.79.

Maia, Lucas Oliveira, Yvan Beaussant, and Ana Cláudia Mesquita Garcia. 2022. “The Therapeutic Potential of Psychedelic-Assisted Therapies for Symptom Control in Patients Diagnosed With Serious Illness: A Systematic Review.” *Journal of Pain and Symptom Management* 63 (6): e725–38. https://doi.org/10.1016/j.jpainsymman.2022.01.024.

McNamee, Sarah, Neşe Devenot, and Meaghan Buisson. 2023. “Studying Harms Is Key to Improving Psychedelic-Assisted Therapy—Participants Call for Changes to Research Landscape.” *JAMA Psychiatry* 80 (5): 411–12. https://doi.org/10.1001/jamapsychiatry.2023.0099.

Mitchell, Jennifer M., Marcela Ot’alora G., Bessel van der Kolk, Scott Shannon, Michael Bogenschutz, Yevgeniy Gelfand, Casey Paleos, et al. 2023. “MDMA-Assisted Therapy for Moderate to Severe PTSD: A Randomized, Placebo-Controlled Phase 3 Trial.” *Nature Medicine* 29 (10): 2473–80. https://doi.org/10.1038/s41591-023-02565-4.

Moreno, Francisco A., Christopher B. Wiegand, E. Keolani Taitano, and Pedro L. Delgado. 2006. “Safety, Tolerability, and Efficacy of Psilocybin in 9 Patients With Obsessive-Compulsive Disorder.” *The Journal of Clinical Psychiatry* 67 (11): 0–0.

Mustafa, Reem A, Brett McQueen, Dmitriy Nikitin, Emily Nhan, Antal Zemplenyi, Michael J DiStefano, Yasmine Kayali, and Marina Richardson. 2024. “Midomafetamine-Assisted Psychotherapy for Post-Traumatic Stress Disorder - Evidence Report.” Institute for Clinical and Economic Review.

Muthukumaraswamy, Suresh D. 2023. “Overcoming Blinding Confounds in Psychedelic Randomized Controlled Trials Using Biomarker Driven Causal Mediation Analysis.” *Expert Review of Clinical Pharmacology* 16 (12): 1163–73. https://doi.org/10.1080/17512433.2023.2279736.

Muthukumaraswamy, Suresh D., Anna Forsyth, and Thomas Lumley. 2021. “Blinding and Expectancy Confounds in Psychedelic Randomised Controlled Trials.” *Expert Review of Clinical Pharmacology*, no. just-accepted.

Nautiyal, Katherine M., and David B. Yaden. 2023. “Does the Trip Matter? Investigating the Role of the Subjective Effects of Psychedelics in Persisting Therapeutic Effects.” *Neuropsychopharmacology* 48 (1): 215–16. https://doi.org/10.1038/s41386-022-01424-z.

Nayak, Sandeep M., Bilal A. Bari, David B. Yaden, Meg J. Spriggs, Fernando E. Rosas, Joseph M. Peill, Bruna Giribaldi, David Erritzoe, David J. Nutt, and Robin Carhart-Harris. 2023. “A Bayesian Reanalysis of a Trial of Psilocybin Versus Escitalopram for Depression.” *Psychedelic Medicine* 1 (1): 18–26. https://doi.org/10.1089/psymed.2022.0002.

Nayak, Sandeep M., Melissa K. Bradley, Bethea A. Kleykamp, Eric C. Strain, Robert H. Dworkin, and Matthew W. Johnson. 2023. “Control Conditions in Randomized Trials of Psychedelics: An ACTTION Systematic Review.” *The Journal of Clinical Psychiatry* 84 (3): 47000. https://doi.org/10.4088/JCP.22r14518.

Nour, Matthew M., Lisa Evans, David Nutt, and Robin L. Carhart-Harris. 2016. “Ego-Dissolution and Psychedelics: Validation of the Ego-Dissolution Inventory (EDI).” *Frontiers in Human Neuroscience* 10:269.

Olson, David E. 2021. “The Subjective Effects of Psychedelics May Not Be Necessary for Their Enduring Therapeutic Effects.” *ACS Pharmacology & Translational Science* 4 (2): 563–67. https://doi.org/10.1021/acsptsci.0c00192.

Olson, Jay A., Léah Suissa-Rocheleau, Michael Lifshitz, Amir Raz, and Samuel P. L. Veissière. 2020. “Tripping on Nothing: Placebo Psychedelics and Contextual Factors.” *Psychopharmacology* 237 (5): 1371–82. https://doi.org/10.1007/s00213-020-05464-5.

Palhano-Fontes, Fernanda, Dayanna Barreto, Heloisa Onias, Katia C. Andrade, Morgana M. Novaes, Jessica A. Pessoa, Sergio A. Mota-Rolim, Flávia L. Osório, Rafael Sanches, and Rafael G. Dos Santos. 2019. “Rapid Antidepressant Effects of the Psychedelic Ayahuasca in Treatment-Resistant Depression: A Randomized Placebo-Controlled Trial.” *Psychological Medicine* 49 (4): 655–63.

Peck, Stephanie Knatz, Samantha Shao, Tessa Gruen, Kevin Yang, Alexandra Babakanian, Julie Trim, Daphna M. Finn, and Walter H. Kaye. 2023. “Psilocybin Therapy for Females with Anorexia Nervosa: A Phase 1, Open-Label Feasibility Study.” *Nature Medicine* 29 (8): 1947–53. https://doi.org/10.1038/s41591-023-02455-9.

Pollan, Michael. 2018. *How to Change Your Mind*. New York: Penguin Random House.

Raison, Charles L., Gerard Sanacora, Joshua Woolley, Keith Heinzerling, Boadie W. Dunlop, Randall T. Brown, Rishi Kakar, et al. 2023. “Single-Dose Psilocybin Treatment for Major Depressive Disorder: A Randomized Clinical Trial.” *JAMA* 330 (9): 843–53. https://doi.org/10.1001/jama.2023.14530.

Rocha, Juliana Mendes, Giordano Novak Rossi, Flávia de Lima Osório, José Carlos Bouso, Gabriela de Oliveira Silveira, Mauricio Yonamine, Alline Cristina Campos, Giuliana

Bertozi, Jaime E. Cecílio Hallak, and Rafael G. Dos Santos. 2021. “Effects of Ayahuasca on the Recognition of Facial Expressions of Emotions in Naive Healthy Volunteers: A Pilot, Proof-of-Concept, Randomized Controlled Trial.” *Journal of Clinical Psychopharmacology* 41 (3): 267–74.

Romeo, B., E. Kervadec, B. Fauvel, L. Strika-Bruneau, A. Amirouche, V. Verroust, P. Piolino, and A. Benyamina. 2024. “Safety and Risk Assessment of Psychedelic Psychotherapy: A Meta-Analysis and Systematic Review.” *Psychiatry Research* 335 (May):115880. https://doi.org/10.1016/j.psychres.2024.115880.

Rorty, Richard. 1979. *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.

Rotz, Robin von, Eva M. Schindowski, Johannes Jungwirth, Anna Schuldt, Nathalie M. Rieser, Katharina Zahoranszky, Erich Seifritz, et al. 2023. “Single-Dose Psilocybin-Assisted Therapy in Major Depressive Disorder: A Placebo-Controlled, Double-Blind, Randomised Clinical Trial.” *eClinicalMedicine* 56 (February). https://doi.org/10.1016/j.eclinm.2022.101809.

Routh, Jonathan C., Douglas R. Bacon, Bradley C. Leibovich, Horst Zincke, Michael L. Blute, and Igor Frank. 2008. “How Long Is Too Long? The Effect of the Duration of Anaesthesia on the Incidence of Non-Urological Complications after Surgery.” *BJU International* 102 (3): 301–4. https://doi.org/10.1111/j.1464-410X.2008.07663.x.

Schenberg, Eduardo Ekman. 2021. “Who Is Blind in Psychedelic Research? Letter to the Editor Regarding: Blinding and Expectancy Confounds in Psychedelic Randomized Controlled Trials.” *Expert Review of Clinical Pharmacology* 14 (10): 1317–19. https://doi.org/10.1080/17512433.2021.1951473.

Schenberg, Eduardo Ekman, Franklin King, João Eusébio da Fonseca, and Leor Roseman. 2024. “Is Poorly Assisted Psilocybin Treatment an Increasing Risk?” *American Journal of Psychiatry* 181 (1): 75–76. https://doi.org/10.1176/appi.ajp.20230664.

Smith, William R., and Dominic Sisti. 2021. “Ethics and Ego Dissolution: The Case of Psilocybin.” *Journal of Medical Ethics* 47 (12): 807–14. https://doi.org/10.1136/medethics-2020-106070.

Strassman, Rick J. 2017. “The Political Correction of Psychedelics. Part 3. Psychedelics as Super-Placebos.” *DMT: The Spirit Molecule* (blog). September 28, 2017. https://www.rickstrassman.com/the-political-correction-of-psychedelics-part-3-psychedelics-as-super-placebos/.

Szigeti, Balázs, and Boris Heifets. 2024. “Expectancy Effects in Psychedelic Trials.” *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, February. https://doi.org/10.1016/j.bpsc.2024.02.004.

Szigeti, Balázs, Brandon Weiss, Fernando E. Rosas, David Erritzoe, David Nutt, and Robin Carhart-Harris. 2024. “Assessing Expectancy and Suggestibility in a Trial of Escitalopram v. Psilocybin for Depression.” *Psychological Medicine*, January, 1–8. https://doi.org/10.1017/S0033291723003653.

Timmermann, Christopher, Leor Roseman, Luke Williams, David Erritzoe, Charlotte Martial, Héléna Cassol, Steven Laureys, David Nutt, and Robin Carhart-Harris. 2018. “DMT Models the Near-Death Experience.” *Frontiers in Psychology* 9:1424.

Villiger, Daniel. 2022. “How Psychedelic-Assisted Treatment Works in the Bayesian Brain.” *Frontiers in Psychiatry* 13. https://www.frontiersin.org/article/10.3389/fpsyt.2022.812180.

———. 2024a. “Giving Consent to the Ineffable.” *Neuroethics* 17 (1): 11. https://doi.org/10.1007/s12152-024-09545-6.

———. 2024b. “Informed Consent Under Ignorance.” *The American Journal of Bioethics* 0 (0): 1–13. https://doi.org/10.1080/15265161.2023.2296429.

Wackermann, Jiří, Marc Wittmann, Felix Hasler, and Franz X. Vollenweider. 2008. “Effects of Varied Doses of Psilocybin on Time Interval Reproduction in Human Subjects.” *Neuroscience Letters* 435 (1): 51–55.

Wampold, Bruce E., and Zac E. Imel. 2015. *The Great Psychotherapy Debate: The Evidence for What Makes Psychotherapy Work*. Second edition. New York: Routledge.

Wampold, Bruce E., Takuya Minami, Sandra Callen Tierney, Thomas W. Baskin, and Kuldhir S. Bhati. 2005. “The Placebo Is Powerful: Estimating Placebo Effects in Medicine and Psychotherapy from Randomized Clinical Trials.” *Journal of Clinical Psychology* 61 (7): 835–54. https://doi.org/10.1002/jclp.20129.

Watts, Rosalind, Camilla Day, Jacob Krzanowski, David Nutt, and Robin Carhart-Harris. 2017. “Patients’ Accounts of Increased ‘Connectedness’ and ‘Acceptance’ after Psilocybin for Treatment-Resistant Depression.” *Journal of Humanistic Psychology* 57 (5): 520–64.

Weil, Andrew. 1972. *The Natural Mind*. Boston: Houghton Mifflin.

Weiss, Brandon, Joshua D. Miller, Nathan T. Carter, and W. Keith Campbell. 2021. “Examining Changes in Personality Following Shamanic Ceremonial Use of Ayahuasca.” *Scientific Reports* 11 (1): 6653. https://doi.org/10.1038/s41598-021-84746-0.

Weiss, Brandon, Leor Roseman, Bruna Giribaldi, David J. Nutt, Robin L. Carhart-Harris, and David Erritzoe. 2024. “Unique Psychological Mechanisms Underlying Psilocybin Therapy Versus Escitalopram Treatment in the Treatment of Major Depressive Disorder.” *International Journal of Mental Health and Addiction*, March. https://doi.org/10.1007/s11469-024-01253-9.

Wen, Alexander, Nikhita Singhal, Brett D.M. Jones, Richard J. Zeifman, Shobha Mehta, Mohammad A. Shenasa, Daniel M. Blumberger, Zafiris J. Daskalakis, and Cory R. Weissman. 2024. “A Systematic Review of Study Design and Placebo Controls in Psychedelic Research.” *Psychedelic Medicine* 2 (1): 15–24. https://doi.org/10.1089/psymed.2023.0028.

Yaden, David B., and Roland R. Griffiths. 2021. “The Subjective Effects of Psychedelics Are Necessary for Their Enduring Therapeutic Effects.” *ACS Pharmacology & Translational Science* 4 (2): 568–72. https://doi.org/10.1021/acsptsci.0c00194.

Yaden, David B., James B. Potash, and Roland R. Griffiths. 2022. “Preparing for the Bursting of the Psychedelic Hype Bubble.” *JAMA Psychiatry* 79 (10): 943–44. https://doi.org/10.1001/jamapsychiatry.2022.2546.