From efficacy to effectiveness: evaluating psychedelic randomised controlled trials for trustworthy evidence-based policy and practice.

Author: Eduardo Ekman Schenberg, PhD

Founder and Director, Instituto Phaneros, São Paulo, Brasil.
Honorary researcher, University College London, UK.
Associate Researcher, project on psychedelics law and regulation (POPLAR), The Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics at Harvard Law School, USA.

Abstract

The recent review of a new drug application for MDMA-assisted therapy for post-traumatic stress disorder by the United States' Food and Drug Administration (FDA) highlighted epistemological and methodological challenges for evidence assessments. Similar challenges will also be faced in reviews of other compounds in early- and late-stage development, like psilocybin for depression.

The regulatory demand for two successful phase 3 randomised controlled trials (RCTs) seems problematic, given a current lack of agreement on what constitutes "success", particularly when psychoactive drug administration is concomitant with (psycho)therapy. These complex arrangements challenge the internal validity of estimated average treatment effect through comparison with conventional control conditions.

This paper reviews the assumptions behind RCTs' current "gold-standard" status in the hierarchy of evidence-based medicine (EBM). Recapitulating known epistemic limits of randomisation and blinding, it emphasises the urgent need to avoid the extrapolation fallacy.

The resulting argument is that the degree of trustworthiness that efficacy — reported in RCTs — will reliably predict effectiveness — in target populations outside RCTs — depends on what type of psychedelic treatments will be regulated. If "stand-alone" drugs for large scale prescription and consumption, trustworthiness should be graded low. On the other hand, for regulation of drug-assisted (psycho)therapies, the degree of trustworthiness can be considered high. The reason being that these two treatment approaches are based on different causal claims with distinct external validities. Therefore, careful assessment of support factors in each is recommended to prevent detrimental consequences, from potential rejection of effective therapies up to medical reversal of eventually approved drugs.

Keywords

Psychedelic, RCT, causal, bias, efficacy, effectiveness, evidence, mechanisms, extrapolation fallacy, medical reversal.

Background and rationale

After being largely dismissed during the last quarter of the 20th century, psychedelics were increasingly reappraised in clinical trials during the first decades of the 21st century. This generated hope amidst an innovation stall in clinical psychopharmacology, concomitant with high unmet medical need[1]. For example, esketamine was approved in many jurisdictions, including the USA and Europe[2], while two phase 3 trials with "MDMA-assisted therapy" for post-traumatic stress disorder (PTSD) were published[3]. Additionally, multiple phase 3 trials are ongoing with psilocybin for the treatment of major[4] and treatment resistant depression[5], and LSD for generalised anxiety disorder is in development[6]. Finally, there is also an increasing number of phase 1 and 2 studies with similar compounds with a wide range of chemical, pharmacological and psychoactive properties[7]. This second biomedical wave, after initial research in the 1950s and early 1960s, unfolded in excessive enthusiasm and hype, especially as a for-profit market of start-ups, fueled by specific investment funds, started claiming intellectual property over these once discarded compounds[8].

---

[1] Langlitz, N. Psychedelic innovations and the crisis of psychopharmacology. BioSoc. 19:37–58, 2024. https://doi.org/10.1057/s41292-022-00294-4.

[2] Mahase E. Esketamine is approved in Europe for treating resistant major depressive disorder. BMJ 2019; 367:l7069 http://doi.org/10.1136/bmj.l7069; Kim, J., Farchione, T., Potter, A., Chen, Q., & Temple, R. (2019). Esketamine for Treatment-Resistant Depression - First FDA-Approved Antidepressant in a New Class. NEKM 381(1), 1–4. https://doi.org/10.1056/NEJMp1903305.

[3] Mitchell, J.M., Bogenschutz, M., Lilienstein, A. et al. MDMA-assisted therapy for severe PTSD: a randomized, double-blind, placebo-controlled phase 3 study. Nat Med 27, 1025–1033 (2021). https://doi.org/10.1038/s41591-021-01336-3; Mitchell, J.M., Ot'alora G., M., van der Kolk, B. et al. MDMA-assisted therapy for moderate to severe PTSD: a randomized, placebo-controlled phase 3 trial. Nat Med 29, 2473–2480 (2023). https://doi.org/10.1038/s41591-023-02565-4.

[4] Phase 3 Study of Psilocybin for Major Depressive Disorder (MDD), https://www.usonainstitute.org/psilocybin;

[5] Compass Pathways initiates UK component of global phase 3 study of psilocybin treatment in treatment-resistant depression, and launches new research center. https://compasspathways.com/compass-pathways-initiates-uk-component-of-global-phase-3-study-of-psilocybin-treatment-in-treatment-resistant-depression-and-launches-new-research-center/.

[6] MindMed Receives FDA Breakthrough Therapy Designation and Announces Positive 12-Week Durability Data From Phase 2B Study of MM120 for Generalized Anxiety Disorder. https://ir.mindmed.co/news-events/press-releases/detail/137/mindmed-receives-fda-breakthrough-therapy-designation-and-announces-positive-12-week-durability-data-from-phase-2b-study-of-mm120-for-generalized-anxiety-disorder.

[7] Psychedelics Drug Development Tracker. https://psychedelicalpha.com/data/psychedelic-drug-development-tracker.

[8] Aday JS, Bloesch EK, Davoli CC. 2019: A year of expansion in psychedelic research, industry, and deregulation. Drug Science, Policy and Law. 2020;6. http://doi.org/10.1177/2050324520974484; Yaden DB, Potash JB, Griffiths RR. Preparing for the Bursting of the Psychedelic Hype Bubble. JAMA Psychiatry. 2022;79(10):943–944. http://doi.org/10.1001/jamapsychiatry.2022.2546.

More recently, the enthusiasm has been balanced by ethical and safety concerns beyond the scope of the current paper[9], which focus on increasing criticism related to efficacy estimates. According to these, the available evidence may be insufficient for securing efficacy and therefore more rigorous studies are being demanded. This perspective seems to have dominated two important events, the “Advancing Psychedelic Clinical Study Design”[10] organised by the Reagan-Udall Foundation for the FDA in January 31st and February 1st, 2024; and the second in April 16th and 17th, 2024, at the European Medicines Agency (EMA), entitled ”EMA Multi-Stakeholder Workshop on Psychedelics — Towards an EU Regulatory Framework”[11]. In both events discussions were centred around safety, efficacy data and how to improve trial designs to account for interactions between (psycho)therapy and psychopharmacology. Finally, a few years after granting MDMA-assisted therapy “breakthrough therapy” status, and following a psychopharmacology advisory committee meeting in June, the United States of America Food and Drug Administration (FDA) did not approve a new drug application (NDA) for MDMA-assisted therapy for PTSD in August 2024, likely based on similar reasons regarding potentially biased efficacy estimates[12].

To better understand these controversies, I first review the criticisms of psychedelic randomised controlled trials (RCTs), and then examine the philosophical assumptions and epistemological premises usually left implicit in these criticisms, and also in regulatory guidance documents. This exercise, carefully conducted and based on philosophy of evidence in medicine, exposes how and why a reliable assessment of the current evidence-base for clinical applications of psychedelics requires consideration of recent conceptual and practical advances in evidence-based medicine (EBM) and evidence-based policy and practice (EBPP). Particularly important is the evolution from the concept of a *hierarchy of evidence* to *evidential pluralism* in EBM+. Consequently, it is necessary to integrate evidence from clinical trials with mechanistic

---

[9] Anderson, B. T., Danforth, A. L., & Grob, C. S. (2020). Psychedelic medicine: safety and ethical concerns. The Lancet Psychiatry, 7(10), 829–830. https://doi.org/10.1016/S2215-0366(20)30146-2; McNamee S, Devenot N, Buisson M. Studying Harms Is Key to Improving PAT—Participants Call for Changes to Research Landscape. JAMA Psychiatry. 2023;80(5):411–412. http://doi.org/10.1001/jamapsychiatry.2023.0099; Marks, M., Cohen, I.G. Psychedelic therapy: a roadmap for wider acceptance and utilization. Nat Med 27, 1669–1671 (2021), https://doi.org/10.1038/s41591-021-01530-3. The Lancet Regional Health-Europe (2023). Psychedelic-assisted psychotherapy: hope and dilemma. The Lancet regional health. Europe, 32, 100727. https://doi.org/10.1016/j.lanepe.2023.100727. McGuire, A. L., Lynch, H. F., Grossman, L. A., & Cohen, I. G. (2023). Pressing regulatory challenges for psychedelic medicine. Science, 380(6643), 347–350. https://doi.org/10.1126/science.adg1324. Smith, W. R., & Appelbaum, P. S. (2022). Novel ethical and policy issues in psychiatric uses of psychedelic substances. Neuropharmacology, 216, 109165. https://doi.org/10.1016/j.neuropharm.2022.109165; McGuire A. et al. Developing an Ethics and Policy Framework for Psychedelic Clinical Care. JAMA Network Open, 2024 (in press).

[10] Advancing Psychedelic Clinical Study Design. Full content available at: https://reaganudall.org/news-and-events/events/advancing-psychedelic-clinical-study-design

[11] EMA multi-stakeholder workshop on psychedelics – Towards an EU regulatory framework. Full content available at https://www.ema.europa.eu/en/events/ema-multi-stakeholder-workshop-psychedelics-towards-eu-regulatory-framework

[12] https://news.lykospbc.com/2024-08-09-Lykos-Therapeutics-Announces-Complete-Response-Letter-for-Midomafetamine-Capsules-for-PTSD

evidence for assessing causal claims, and to reliably extrapolate results from trial samples to larger and different target populations.

This approach reveals that the two distinct treatment models currently advocated for — psychedelic-assisted (psycho)therapy (usually referred to as PAT or PAP) on one end, and psychedelic drugs, on the other — make two different causal claims. The reliability of currently published RCTs and their assessment as sufficient or not for drug approvals depend on the purposes of each and the conditional epistemic premises they rely on.

Review of current criticism

With a few early papers[13], criticism regarding efficacy estimates from psychedelic RCTs has increased since 2020. The most frequent concern regards unblinding (or unmasking)[14] of treatment allocation, and how expectancy can bias results[15]. This is because participants in experimental groups almost certainly realise they received a potent psychoactive compound, and therefore have high — or positive — expectancy; while participants in the control groups most likely suspect they did not receive the drug, and may therefore be frustrated and have low — or negative — expectancy[16].

---

[13] Sellers, E. M., & Leiderman, D. B. (2018). Psychedelic Drugs as Therapeutics: No Illusions About the Challenges. Clin. Pharm. Ther., 103(4), 561–564. https://doi.org/10.1002/cpt.776; Barnby, J. M., & Mehta, M. A. (2018). Psilocybin and Mental Health-Don't Lose Control. Frontiers in Psychiatry, 9, 293. https://doi.org/10.3389/fpsyt.2018.00293.

[14] In spite of important calls to avoid the "blinding" terminology in respect for people with disabilities, I opted to keep with the most widely used term, and also to avoid confusion with the problem of counteracting mechanisms known as the problem of "masking" (Illari, P. M. (2011) 'Mechanistic Evidence: Disambiguating the Russo–Williamson Thesis', Int. Stud. Philos. Sci., 25(2), pp. 139–157. http://doi.org/10.1080/02698595.2011.574856.

[15] For the most recent review, see Szigeti, B., & Heifets, B. D. (2024). Expectancy Effects in Psychedelic Trials. Biological psychiatry. Cognitive neuroscience and neuroimaging, S2451-9022(24)00055-7. Advance online publication. https://doi.org/10.1016/j.bpsc.2024.02.004.

[16] Burke, M.J., Blumberger, D.M. Caution at psychiatry's psychedelic frontier. Nat Med 27, 1687–1688 (2021). https://doi.org/10.1038/s41591-021-01524-1; Halvorsen, J.Ø., Naudet, F. & Cristea, I.A. Challenges with benchmarking of MDMA-assisted psychotherapy. Nat Med 27, 1689–1690 (2021). https://doi.org/10.1038/s41591-021-01525-0; Muthukumaraswamy S, Forsyth A, Sumner RL. The challenges ahead for psychedelic 'medicine'. Aust. NZ. J. of Psychiatry. 2022;56(11):1378-1383. https://doi:10.1177/00048674221081763; Aday, J. S., Heifets, B. D., Pratscher, S. D., Bradley, E., Rosen, R., & Woolley, J. D. (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; van Elk M, Fried EI. History repeating: guidelines to address common problems in psychedelic science. Ther. Adv. Psychopharm. 2023;13. https://doi:10.1177/20451253231198466; Volkow ND, Gordon JA, Wargo EM. Psychedelics as Therapeutics—Potential and Challenges. JAMA Psychiatry. 2023;80(10):979–980. https://doi:10.1001/jamapsychiatry.2023.1968; McClure-Begley, T.D., Roth, B.L. The promises and perils of psychedelic pharmacology for psychiatry. Nat Rev Drug Discov 21, 463–473 (2022). https://doi.org/10.1038/s41573-022-00421-7; Hovmand OR, Poulsen ED, Arnfred S, Storebø OJ. Risk of bias in randomized clinical trials on psychedelic medicine: A systematic review. J. Psychopharmacol. 2023;37(7):649-659. https://doi:10.1177/02698811231180276; Soliman, P. S., Curley, D. E., Capone, C., Eaton, E., & Haass-Koffler, C. L. (2024). In the new era of psychedelic assisted therapy: A systematic review of study methodology in randomized controlled trials. Psychopharmacology, 10.1007/s00213-024-06598-6. Advance online publication. https://doi.org/10.1007/s00213-024-06598-6; Alexander Wen, Nikhita Singhal, Brett D.M. Jones, Richard J. Zeifman, Shobha Mehta, Mohammad A.

The most comprehensive critique was presented by Prof. Suresh Muthukumaraswamy in the University of Auckland, New Zealand. The main argument is that biases can happen in RCTs throughout planning, execution, analysis, interpretation of data and publication of results. Most of these are neither new nor specific to psychedelics, but are more likely to happen with potent psychoactive drugs whose effects are quite extraordinary and unmistakable by those who experience them, and also to observers. This is due to the drugs' characteristic psychoactive effects likely never going unnoticed by people who take them (especially in medium to high doses), and also by research staff due to characteristic overt (and covert) emotional and behavioural manifestations. A non-exhaustive list includes profusely smiling, laughing, screaming, crying and sobbing, as well as unusual and wide range of movements, bodily postures and facial expressions, etc. The main worry is that expectancy biases can, and likely did, inflate estimates of average treatment effects (ATEs), threatening the reliability of causal inferences. This would thus risk spurious correlations being mistakenly treated as causation: "It is theoretically possible that the entire estimated effect size of these interventions could be accounted for by these confounds which would mean that they provide no evidence for efficacy"[17].

From hierarchy to evidential pluralism

Implicit in these criticisms is the larger context of EBM, a so-called "paradigm" which emerged during the 1970s and 80s, gaining prominence in medicine during the last two decades of the 20th century[18], precisely when therapeutic uses of psychedelics were largely under-researched. According to the highly influential principles of EBM, RCTs are considered "the gold standard" and are placed at the top of a hierarchy of evidence. Above RCTs there are other methods which strictly depend on them, such as systematic reviews and meta-analyses. In EBM's parlance, RCTs are the only method capable of gathering unbiased evidence, and therefore policy decisions should be based, if not exclusively, predominantly on them, to the

---

Shenasa, Daniel M. Blumberger, Zafiris J. Daskalakis, and Cory R. Weissman. Risk of bias in randomized clinical trials on psychedelic medicine: A systematic review. Psychedelic Medicine;2(1), https://doi.org/10.1089/psymed.2023.0028; Nayak, S. M., Bradley, M. K., Kleykamp, B. A., Strain, E. C., Dworkin, R. H., & Johnson, M. W. (2023). Control Conditions in Randomized Trials of Psychedelics: An ACTTION Systematic Review. J. Clin. Psych., 84(3), 22r14518. https://doi.org/10.4088/JCP.22r14518; Bahji, A., Lunsky, I., Gutierrez, G., & Vazquez, G. (2023). Efficacy and Safety of Four Psychedelic-Assisted Therapies for Adults with Symptoms of Depression, Anxiety, and Posttraumatic Stress Disorder: A Systematic Review and Meta-Analysis. J. Psych. Drugs, 1–16. https://doi:10.1080/02791072.2023.2278586; Hall WD, Humphreys K. Is good science leading the way in the therapeutic use of psychedelic drugs? Psychological Medicine. 2022;52(14):2849-2851. https://doi:10.1017/S0033291722003191; Noorani, T., Bedi, G., & Muthukumaraswamy, S. (2023). Dark loops: contagion effects, consistency and chemosocial matrices in PAT trials. Psyc. Med, 53(13):5892–5901. https://doi.org/10.1017/S0033291723001289.

[17] Muthukumaraswamy, S. D., Forsyth, A., & Lumley, T. (2021). Blinding and expectancy confounds in psychedelic randomized controlled trials. Expert Rev. Clin. Pharmacol., 14(9), 1133–1152. https://doi.org/10.1080/17512433.2021.1933434; Muthukumaraswamy S. D. (2023). Overcoming blinding confounds in psychedelic randomized controlled trials using biomarker driven causal mediation analysis. Expert Rev. Clin. Pharmacol., 16(12), 1163–1173. https://doi.org/10.1080/17512433.2023.2279736.

[18] Howick, J. THE PHILOSOPHY OF EVIDENCE-BASED MEDICINE (Wiley, 2011). http://doi.org/10.1002/9781444342673.

neglect of all other sources of evidence. This view has dominated for decades, as the following examples illustrate:

“The only source of reliable evidence about the usefulness of almost any sort of therapy […] is that obtained from well-planned and carefully conducted randomised [...] clinical trials” Tukey, 1977[19]

“Because the randomised trial, and especially the systematic review of several randomised trials, is so much more likely to inform us and so much less likely to mislead us, it has become the ‘gold standard’ ” Sackett, 1996[20]

“The strength of an RCT is that [...] the study’s design controls for all the factors, other than the planned intervention, that might lead to different outcomes. This is terrifically powerful.”[21] Prasad and Cifu, 2019.

Claims of this kind originate from common misunderstandings about RCTs’ actual epistemic strengths[22]. A general and widespread misconception is that ATEs ensure causal inference. In pharmacology this is usually misinterpreted as proof that the drug is the cause of the ATE in the trial sample because all other factors were supposedly controlled by random allocation and blinding. However, this is not so. The high internal validity of RCTs is defined as “the validity of inferences about whether observed covariation ... reflects a causal relationship”[23]. This is what gives RCTs their epistemic power, as the method is deductive. That is, if all the premises are met, the conclusion logically follows from the results, in the ideal scenario[24]. However, in practice, there are many premises that must be met for RCTs to actually approximate the ideal case. For example, Cook listed 26 assumptions that should be met for RCTs to grant causal inference. As it is very rare, or never the case, that all are met for any single trial, or even a small collection of trials, he concluded that “it is much easier to argue that RE [randomised experiments] represents a high standard of causal inference than to argue that it is the gold standard whose use guarantees perfectly valid internal validity.”[25].

The first reason for this is because balance through random allocation (not to be confused with random sampling from a population) is not actually granted. Despite being frequently said to guarantee that the comparison groups in a trial are (perfectly) balanced at

---

[19] Apud Worral, J. Evidence and Ethics in Medicine. Perspect. Biol. Med., 51(3):418-431, 2008. https://doi:10.1353/pbm.0.0040.
[20] Apud Goldenberg, MJ. "Iconoclast or Creed?: Objectivism, Pragmatism, and the Hierarchy of Evidence." Perspect. Biol. Med, 52(2):168-187, 2009. https://doi.org/10.1353/pbm.0.0080.
[21] Cifu, AS. and VK. Prasad. ENDING MEDICAL REVERSAL: Improving Outcomes, Saving Lives (Johns Hopkins University Press, 2015). https://doi.org/10.1353/book.49286.
[22] Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. SSM, 210:2–21. https://doi.org/10.1016/j.socscimed.2017.12.005.
[23] Shadish, W.R., Cook, T.D., Campbell, D.T. EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS FOR GENERALIZED CAUSAL INFERENCE (Houghton Mifflin, Boston, MA, 2002)
[24] Cartwright, N. Are RCTs the Gold Standard?. BioSoc. 2:11–20 (2007). https://doi.org/10.1017/S1745855207005029.
[25] Cook TD. (2018). Twenty-six assumptions that have to be met if single random assignment experiments are to warrant "gold standard" status: A commentary on Deaton and Cartwright. SSM, 210:37–40. https://doi.org/10.1016/j.socscimed.2018.04.031.

baseline, and therefore equivalent except for the variable under study, single random treatment allocation cannot actually grant unbalanced groups to all variables. By the simple laws of probability, balance would only be achieved with repeated randomisations, indefinitely. Medical research does not, and cannot, indefinitely repeat experiments, and therefore single randomised experiments, no matter how rigorous, are never entirely free of biases. The larger the number of potential confounders, the worse it gets. This is sometimes referred to as “unhappy randomisation”, and re-randomisation based on post-randomisation verification of confounders is often promoted as a remedy. However, only known confounders can be checked for post-randomisation, but unknown confounders must never be ruled out. Therefore, RCTs can always be biassed in principle, and frequently are in practice[26]. As Nobel Prize winner (2015) Angus Deaton and Nancy Cartwright, recipient of the Martin R. Lebowitz Prize for Philosophical Achievement (2017) and the Carl Gustav Hempel Award (2018) put it: “If we were to repeat the trial many times, the over-representation of the unbalanced causes will sometimes be in the treatments and sometimes in the controls. The imbalance will vary over replications of the trial, and although we cannot see this from our single trial, we should be able to capture its effects on our estimate of the ATE from an estimated standard error. This was Fisher's insight: not that randomization balanced covariates between treatments and controls but that, conditional on the caveat that no post-randomization correlation with covariates occurs, randomization provides the basis for calculating the size of the error. Getting the standard error and associated significance statements right are of the greatest importance; therein lies the virtue of randomization, not that it yields precise estimates through balance.”[27] To summarise, the common assumption that RCTs yield unbiased results through balanced groups is wrong, and it would be just through many repetitions that randomisation would approach unbiasedness.

The second reason for imperfect internal validity is the caveat that after randomisation imbalances can occur, and therefore other remedies may be necessary. That is what blinding aims at: controlling for post-randomisation confounders. However, contrary to randomisation, blinding has a poorly studied history, which led Ted Kaptchuk to conclude that “the aura of objectivity and neutrality attached to blind assessment itself may have benefited from this absence of a past”[28]. Furthermore, empirical evidence shows that blinding fails much more often than generally accepted and understood. For example, researchers searching Medline and Cochrane found 90 reports attempting to assess blinding, of which 58 (62%) specifically assessed participants blinding. They found inconsistencies in the methods and timing of assessments, concluding that blinding was questionable[29]. Another examination inspected the ten most cited RCTs in biomedicine (each with >6500 citations at the time of analysis), revealing that only two (20%) were successfully blinded[30]. In a sample of 200 randomly selected RCTs in

---

[26] Worrall, J. “Why There’s No Cause to Randomize.” Brit. J. Phil. Sci., 58(3):pp. 451–88, 2007. http://www.jstor.org/stable/30115184.

[27] Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. SSM, 210:2–21. https://doi.org/10.1016/j.socscimed.2017.12.005.

[28] Kaptchuk, TJ. “Intentional Ignorance: A History of Blind Assessment and Placebo Controls in Medicine.” Bull. Hist. Med., 72(3):389–433, 1998. http://www.jstor.org/stable/44445075.

[29] Boutron, I., Guittet, L., Estellat, C., Moher, D., Hróbjartsson, A., & Ravaud, P. (2007). Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. PLoS Med, 4(2):e61. https://doi.org/10.1371/journal.pmed.0040061.

[30] Krauss A. (2018). Why all randomised controlled trials produce biased results. Ann. Med., 50(4):312–322. https://doi.org/10.1080/07853890.2018.1453233

five of the top medical and four of the top psychiatric journals, only 7% and 9%, respectively, were found to report blinding assessments, while blinding had been lost in 60% of those who tested for it[31]. Similarly, from a random sample of 300 studies from PubMed, only 24 (8%) explicitly reported risk of unblinding[32]. In a border sample of the literature, randomly selecting 1599 studies, reporting of blinding assessments was found between 2% and 8%[33], while in the hundreds of antidepressant trials in the last two decades it was between 5% and 7%[34]. Philosopher John Worral thus seems correct and precise when concluding that blinding "is an ideal more often honoured in the breach than the observance"[35]. This is not to say that the blinding methodology and the confounders it aims to control for do not matter, but if the method frequently fails, besides poor practices it is possible that the underlying epistemic premises are weak. In the case of psychopharmacology and psychiatry, it seems clearly unreasonable to expect people to experience improvements in subjective feelings, especially through the use of psychoactive substances, without becoming consciously aware of these changes in themselves. Furthermore, as blinding became paramount to claims of validity (or lack thereof), its epistemic premises became pervasively misunderstood. As statistician Stephen Senn put it: "The whole point of a successful double blind trial is that there should be unblinding through efficacy. [...] If the treatment is effective the guesses will distribute unequally between the arms of the trial and the trial will then be declared 'not blind'"[36]. Therefore, the usual calls to invalidate studies due to unblinding suffer not only from the fact that reporting unblinding in biomedicine is rare (<10%), but that demanding studies to be unblinded from beginning to end is epistemically flawed, particularly in cases of subjective symptoms. In the case of psychiatry and psychopharmacology, this is because the conditions are subjectively felt and because drugs specifically modulate functions of the central nervous system including interoception, proprioception, self-awareness and consciousness. In these cases it is important not only to measure and report blinding assessments and rates of unblinding, but to clearly distinguish between malicious and benign unblinding, which are related to side effects and efficacy, respectively[37].

---

[31] Fergusson, D., Glass, K. C., Waring, D., & Shapiro, S. (2004). Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials. BMJ, 328(7437):432. https://doi.org/10.1136/bmj.37952.631667.EE.

[32] Bello, S., Moustgaard, H., & Hróbjartsson, A. (2014). The risk of unblinding was infrequently and incompletely reported in 300 randomized clinical trial publications. J. Clin. Epidemiol., 67(10):1059–1069. https://doi.org/10.1016/j.jclinepi.2014.05.007.

[33] A Hróbjartsson, E Forfang, MT Haahr, B Als-Nielsen, S Brorson, Blinded trials taken to the test: an analysis of randomized clinical trials that report tests for the success of blinding, Int. J. Epidemiol., 36(3):654–663, 2007. https://doi.org/10.1093/ije/dym020.

[34] Lin, Y. H., Sahker, E., Shinohara, K., Horinouchi, N., Ito, M., Lelliott, M., Cipriani, A., Tomlinson, A., Baethge, C., & Furukawa, T. A. (2022). Assessment of blinding in randomized controlled trials of antidepressants for depressive disorders 2000-2020: A systematic review and meta-analysis. E. Clin. Med., 50, 101505. https://doi.org/10.1016/j.eclinm.2022.101505

[35] Worrall, J. (2022). Philosophy of Science Meets Medicine (Again): A Clearer-Sighted View of the Virtues of Blinding and of Tests for Blinding in Clinical Trials. In: Gonzalez, W.J. (eds) CURRENT TRENDS IN PHILOSOPHY OF SCIENCE (Synthese Library, vol 462. Springer, Cham). https://doi.org/10.1007/978-3-031-01315-7_2.

[36] Senn S. J. (2004). Turning a blind eye: authors have blinkered view of blinding. BMJ, 328(7448):1135–1136. https://doi.org/10.1136/bmj.328.7448.1135-b.

[37] Schenberg E. E. (2021). Who is blind in psychedelic research? Letter to the editor regarding: blinding and expectancy confounds in psychedelic randomized controlled trials. Exp. Rev. Clin. Pharmacol.,

Despite their important strengths, the weaknesses of randomisation and blinding undermine the frequent overconfidence in RCTs' internal validity. And this is furthermore conflated with a second, and more widespread, problem: the assumption that an unbiased ATE estimate in a RCT, being causally determined due to supposed control of all other confounders, grants exportability or generalizability of this cause beyond the RCT[38]. This is the problem of external validity, defined as "the validity of inferences about whether the cause-effect relationship holds over variation in persons, settings, treatment variables, and measurement variables"[39]. While internal validity refers to the degree of certainty that any outcome was causally determined, external validity refers to the degree of confidence that similar outcomes may be obtained elsewhere, in different people, historical times and/or circumstances (or the "truthfulness of the result" and the "applicability of the results")[40]. As Cartwright put it: "For policy and practice we do not need to know 'it works somewhere'. We need evidence for 'it-will-work-for-us' claims: the treatment will produce the desired outcome in our situation as implemented there. How can we get from it-works-somewhere to it-will-work-for-us?"[41]. In EBM and policy making based on it, it is quite common for internal validity to be explicitly prioritised while external validity is disregarded: "researchers, funding agencies, ethics committees, the pharmaceutical industry, medical journals, and governmental regulators alike all neglect external validity"[42].

Discernment between these two concepts requires a careful distinction between *efficacy*, which is measured in RCTs, and *effectiveness*, or what needs to be estimated for populations beyond trial samples[43]. It is effectiveness, then, that matters most for EBPP. To quote Deaton and Cartwright again: "The literature discussing RCTs has paid more attention to obtaining results than to considering what can justifiably be done with them". However, these points seem to be missing in the US FDA's draft guidance "Psychedelic Drugs: Considerations for Clinical Investigations Guidance for Industry"[44], where "efficacy" appears 5 times and "effectiveness" 6 times, without clear definitions; and also in the EMA's draft "Guideline on clinical investigation of

14(10):1317–1319. https://doi.org/10.1080/17512433.2021.1951473; Schenberg E. E. (2024). Psychedelic skepticism: back to the sixties?. Ther. Adv. Psychopharmacol., vol 14. https://doi.org/10.1177/20451253241243242.

[38] Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. SSM, 210:2–21. https://doi.org/10.1016/j.socscimed.2017.12.005.

[39] Shadish, W.R., Cook, T.D., Campbell, D.T. EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS FOR GENERALIZED CAUSAL INFERENCE (Houghton Mifflin, Boston, MA, 2002).

[40] Ralph I. Horwitz, Burton Singer, Introduction. What works? And for whom?, SSM, 210: 22-25, 2018. https://doi.org/10.1016/j.socscimed.2018.05.013.

[41] Cartwright N. (2011). A philosopher's view of the long road from RCTs to effectiveness. Lancet, 377(9775):1400–1401. https://doi.org/10.1016/s0140-6736(11)60563-1.

[42] Rothwell P. M. (2005). External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet, 365(9453):82–93. https://doi.org/10.1016/S0140-6736(04)17670-8.

[43] Cartwright N. What Is This Thing Called "Efficacy"? In: Mantzavinos C, ed. PHILOSOPHY OF THE SOCIAL SCIENCES: PHILOSOPHICAL THEORY AND SCIENTIFIC PRACTICE (Cambridge University Press, 2009). https://doi.org/10.1017/CBO9780511812880.016.

[44] FDA, 2023. Psychedelic Drugs: Considerations for Clinical Investigations. Avaliable at https://www.fda.gov/regulatory-information/search-fda-guidance-documents/psychedelic-drugs-considerations-clinical-investigations (last accessed May, 1st, 2024).

medicinal products in the treatment of depression"[45], which briefly mentions psychedelic drugs, using "efficacy" 60 times while never employing the term "effectiveness".

The assumption that efficacy unproblematically translates to effectiveness, that is, that efficacy is generalizable to other patient populations in other settings and contexts, with different age, gender, race, comorbidities, cultures and socioeconomic status is known as fallacy of free extrapolation, extrapolator's fallacy or fallacy of simple extrapolation[46]. This is intrinsically related to the policy challenges of how to appropriately use results from RCTs[47], and to the distinction of facts in particular situations from general truths[48]. Historically tragic examples of the extrapolation fallacy are the pervasiveness of biomedical studies with mostly male and white participants, which persists despite repeated calls from feminist and racial movements for decades[49]. RCTs with only white men have balanced groups regarding gender and ethnicity and can therefore have high internal validity, but do little or nothing to grant extrapolation of results to women and people of colour. As Jon Williamson observed: "Any adequate causal epistemology needs to explain how extrapolation is possible and needs to clarify the logic of extrapolation."[50]

Thus, RCTs are undeniably important methods in evidence assessments, especially for simple interventions where it is especially well suited, such as historical cases including streptomycin for tuberculosis, the polio vaccine and treatments for rheumatic fever[51]. But RCTs also have important epistemic limitations, especially in prolonged and complex interventions[52]. The first limitation is to detect harms, given that sample sizes calculated to assess efficacy (usually in the range of hundreds to thousands of patients) are usually too low to reliably and consistently detect rare and very rare adverse events (one in a thousand or one in hundreds of thousands, respectively). Secondly, because of the tradeoff between internal and external validity. Despite their ascendance to, and consolidation at, the top of EBM's rigid hierarchy, these limitations were clearly recognized by the early pioneers and proponents of RCTs:

"Any belief that the controlled trial is the only way would mean not that the pendulum had swung too far but that it had come right off the hook"

---

45 https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-clinical-investigation-medicinal-products-treatment-depression-revision-3_en.pdf

46 Riegelman R. (1979). The fallacy of free extrapolation. Postgrad. Med., 66(3), 189–194. https://doi.org/10.1080/00325481.1979.11715257; Williamson, J. (2019) Establishing Causal Claims in Medicine, Int. Stud. Philos. Sci., 32(1):33-61, http://doi.org/10.1080/02698595.2019.1630927; Fuller, J. The myth and fallacy of simple extrapolation in medicine. Synthese, 198:2919–2939 (2021). https://doi.org/10.1007/s11229-019-02255-0.

47 Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. SSM, 210:2–21. https://doi.org/10.1016/j.socscimed.2017.12.005.

48 Ashcroft R. E. (2004). Current epistemological problems in evidence based medicine. J. Med. Ethics, 30(2):131–135.

49 MM. Lock, V-K. Nguyen. AN ANTHROPOLOGY OF BIOMEDICINE (Wiley, 2nd Edition, 2018). ISBN: 978-1-119-06913-3.

50 Jon Williamson (2019) Establishing Causal Claims in Medicine, Int. Stud. Phil. Sci., 32(1):33-61. http://doi.org/10.1080/02698595.2019.1630927.

51 Solomon, M. Just a paradigm: evidence-based medicine in epistemological context. Euro. J. Phil. Sci., 451 (2011). https://doi.org/10.1007/s13194-011-0034-6;

52 Victora, C. G., Habicht, J. P., & Bryce, J. (2004). Evidence-based public health: moving beyond randomized trials. Am. J Pub. Health, 94(3):400–405. https://doi.org/10.2105/ajph.94.3.400.

Sir August Bradford Hill, 1966[53]

"Between measurements based on RCTs and benefit … in the community there is a gulf which has been much under-estimated"

A L Cochrane, 1971[54]

These epistemic limitations undermine common practices of relying solely on RCTs for causal inference, and led to serious challenges to their "gold standard" status and epistemic authority in evidence hierarchies and policy deliberations[55]. These unfolded in the EBM+ movement (www.ebmplus.org), grounded in the Russo-Williamson Thesis (RWT) and evidential pluralism: "the thesis that one typically needs both evidence of correlation and evidence of mechanisms to establish a causal claim"[56]. This important development proposes that causal assessments in medicine require not only evidence of statistical correlations from (randomised) experiments, but also mechanistic evidence. In this expanded epistemology, making all evidence explicit and being explicit about their use in evidence evaluations is essential, and this includes evidence of mechanisms.

Towards reliable EBPP for psychedelic therapies

The incorporation of evidence of mechanisms into medical assessments is contentious, given many known cases of problems in treatment implementation and scaling up because of flawed mechanistic reasoning, which were central to the establishment of the hierarchy of EBM

---

[53] Apud Concato, J., & Horwitz, R. I. (2018). Randomized trials and evidence in medicine: A commentary on Deaton and Cartwright. SSM, 210:32–36. https://doi.org/10.1016/j.socscimed.2018.04.010.

[54] Apud Rothwell P. M. (2005). External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet, 365(9453):82–93. https://doi.org/10.1016/S0140-6736(04)17670-8.

[55] Goldenberg M. J. (2006). On evidence and evidence-based medicine: lessons from the philosophy of science. SSM, 62(11): 2621–2632. https://doi.org/10.1016/j.socscimed.2005.11.031; Ashcroft R. E. (2004). Current epistemological problems in evidence based medicine. J. Med. Ethics, 30(2):131–135. https://doi.org/10.1136/jme.2003.007039; Solomon, M. Just a paradigm: evidence-based medicine in epistemological context. Eur. J. Phil. Sci., 451 (2011). https://doi.org/10.1007/s13194-011-0034-6; Cartwright, N. Are RCTs the Gold Standard?. BioSoc. 2, 11–20 (2007). https://doi.org/10.1017/S1745855207005029; Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. SSM, 210:2–21. https://doi.org/10.1016/j.socscimed.2017.12.005; Cook T. D. (2018). Twenty-six assumptions that have to be met if single random assignment experiments are to warrant "gold standard" status: A commentary on Deaton and Cartwright. SSM, 210:37–40. https://doi.org/10.1016/j.socscimed.2018.04.031; Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2013). The evidence that evidence-based medicine omits. Prev. Med., 57(6):745–747. https://doi.org/10.1016/j.ypmed.2012.10.020; Clarke, B., Gillies, D., Illari, P. et al. Mechanisms and the Evidence Hierarchy. Topoi, 33:339–360 (2014). https://doi.org/10.1007/s11245-013-9220-9; Frieden T. R. (2017). Evidence for Health Decision Making - Beyond Randomized, Controlled Trials. NEJM, 377(5), 465–475. https://doi.org/10.1056/NEJMra1614394.

[56] Russo, F, & Williamson, J. (2007) Interpreting Causality in the Health Sciences, Int. Stud. Philos. Sci., 21(2):157-170, https://doi.org/10.1080/02698590701498084; V.-P. Parkkinen et al., EVALUATING EVIDENCE OF MECHANISMS IN MEDICINE (SpringerBriefs in Philosophy), https://doi.org/10.1007/978-3-319-94610-8_1; for review and response to criticism, see also Williamson, J. (2019) Establishing Causal Claims in Medicine, Int. Stud. Phil. Sci., 32(1):33-61, https://doi.org/10.1080/02698595.2019.1630927.

privileging RCTs and leaving mechanistic reasoning at the bottom[57]. However, the difference between mechanistic reasoning and mechanistic evidence matters. As Miriam Solomon summarised it: “Incompleteness is the consequence of there being [...] complex interaction of multiple mechanisms in a chaotic and multidimensional system. There are possible hidden mechanisms everywhere in mechanistic stories, despite an easy impression of narrative or causal completeness”[58]. Extra care should also be taken with different meanings commonly attributed to mechanisms in neuroscience, with the distinction between reductive and non-reductive mechanisms being particularly important. In the non-reductive sense, mechanisms are to be found not only in the micro-level phenomena of molecular and cellular biology, but can also include behavioural, psychological, relational and social factors[59].

Therefore, to avoid repeating pre-EBM mistakes, mechanistic reasoning must not be confused with mechanistic evidence. As these can come from many sources, from basic *in vitro* and *in vivo* experiments up to human studies, the reliability of mechanistic evidence assessment requires careful procedures already available[60,61]. It is thus imperative to not entirely accept mechanistic reasoning as well as to not entirely dismiss mechanistic evidence, at the risk of remaining “blind to mechanisms of explanation and causation”[62]. Although relatively recent, evidential pluralism has roots in the nine famous Bradford Hill’s criteria for assessing causal relationships, almost half of which are evidence of statistical associations and the other half are evidence of mechanisms[63].

Thus to safely extrapolate, Cartwright and Hardie proposed working with causal principles and support factors[53]. For any causal inference in RCTs there are support factors which make that causal connection work in the study sample with a complex causal structure in place, which can’t be overlooked because of assumptions related to randomisation and blinding. To reliably extrapolate, it is not only necessary to ascertain that the study sample represents the target population, but also that they have similar causal structures. In other words, that the indispensable support factors will be present and that no overriding derailing factors (eg improper touch, judgmental attitudes or prejudices from therapists, lack of family support, etc) will not intervene to prevent the cause obtaining in the target population. A trustworthy causal

---

[57] Howick, J. THE PHILOSOPHY OF EVIDENCE-BASED MEDICINE (Wiley, 2011). http://doi.org/10.1002/9781444342673; Williamson, J. (2019) Establishing Causal Claims in Medicine, Int. Stud. Phil. Sci., 32(1):33-61. https://doi.org/10.1080/02698595.2019.1630927.

[58] Solomon, M. MAKING MEDICAL KNOWLEDGE (Oxford, 2015). https://doi.org/10.1093/acprof:oso/9780198732617.001.0001.

[59] Ross, L.N., Bassett, D.S. Causation in neuroscience: keeping mechanism meaningful. Nat. Rev. Neurosci. 25, 81–90 (2024). https://doi.org/10.1038/s41583-023-00778-7

[60] Cartwright, N. and Hardie, J. EVIDENCE-BASED POLICY: A PRACTICAL GUIDE TO DOING IT BETTER (New York, 2012), https://doi.org/10.1093/acprof:osobl/9780199841608.001.0001;

[61] Parkkinen VP, Wallmann C, Wilde M, et al. EVALUATING EVIDENCE OF MECHANISMS IN MEDICINE: PRINCIPLES AND PROCEDURES (Cham: Springer, 2018). http://doi.org/10.1007/978-3-319-94610-8; Pérez-González S. (2024). Evidence of mechanisms in evidence-based policy. Stud. His. Phil. Sci., 103:95–104. https://doi.org/10.1016/j.shpsa.2023.11.006.

[62] Ashcroft R. E. (2004). Current epistemological problems in evidence based medicine. J. Med. Ethics, 30(2):131–135. https://doi.org/10.1136/jme.2003.007039.

[63] Hill AB. (1965). The Environment And Disease: Association Or Causation?. Proc. Roy. Soc. Med., 58(5):295–300; Worrall, J. (2011). Causality in medicine: getting back to the Hill top. Prevent. Med., 53(4-5):235–238. https://doi.org/10.1016/j.ypmed.2011.08.009; Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2013). The evidence that evidence-based medicine omits. Prevent. Med., 57(6):745–747. https://doi.org/10.1016/j.ypmed.2012.10.020.

principle is essential to obtaining good arguments that what worked “somewhere” (in RCTs) will also work “here”, at each instance where that causal principle is being considered for implementation and scaling up in diverse target populations.

A Causal Principle was formalised as:

$$CP : y(i)c = a_1 + a_2 y_0(i) + a_3 b(i)x(i) + a_4 z(i)$$[64]

where CP stands for Causal Principle, y is the outcome of interest (with c denoting caused by), x is the variable under study, b are support factors which interact with x, z are support factors independent of x, and a’s are constants across individuals. Therefore, for each individual i, a causal outcome y(i) is the sum of a base level, y0(i), with the variable being studied, x(i), modulated by all possible interacting support factors, *b(i)*, plus additional effects from other support factors independent from x, z(i), each term in the equation having its relevant constant modulators, a1, a2, a3 and a4, which can be considered boost factors. For simplicity only sums are shown, but non-linear effects are possible.

The concept of support factors is essential, because in biology and medicine it is actually rare that diseases and treatments are caused by the classic deterministic description of causality as “same cause, same effect”,[65]. However RCTs are particularly poorly suited at identifying and studying support factors[66].

In the therapeutic applications of psychedelics, establishing reliable EBPP for subsequent treatment implementation in diverse healthcare systems in different jurisdictions, for different populations with different characteristics than those in RCTs, requires attention to support factors. Moreover, the complex nature of the human psychedelic experience and the inter-relational processes that go on between patients and therapists (and research staff) during RCTs should not be dismissed.

It is clear that there is a spectrum of different approaches to therapeutic uses of psychedelics, with stand-alone drug administration on one extreme and PAT on the other end. These are not just two different clinical practices, but two distinct conceptual frameworks making different causal claims. Even if the hypothesis is that the drug only is the causal mechanism, support factors play roles. For psychiatric disorders these are likely more relevant, and PAT may be an extreme case where support factors are paramount. As Deaton and Cartwright remarked: “whether, and in what ways, an RCT result is evidence depends on exactly what the hypothesis is for which the result is supposed to be evidence, and that what kinds of hypotheses these will be depends on the purposes to be served”[66].

In PAT the drug is one factor among many others related to efficacy. In such conceptualization, a causal principle would be contemplated by the full formalisation of Cartwright and Hardie. Here the outcome might be changes measured through scores in

---

[64] Cartwright, N., and Hardie, J. EVIDENCE-BASED POLICY: A PRACTICAL GUIDE TO DOING IT BETTER (New York, 2012), https://doi.org/10.1093/acprof:osobl/9780199841608.001.0001;

[65] Rocca, E.; Anjum, R.L. Causal Evidence and Dispositions in Medicine and Public Health. Int. J. Environ. Res. Public Health 2020, 17, 1813. https://doi.org/10.3390/ijerph17061813; Rani Lill Anjum, Samantha Copeland, Elena Rocca (Editors). RETHINKING CAUSALITY, COMPLEXITY AND EVIDENCE FOR THE UNIQUE PATIENT (Springer Cham, 2020). https://doi.org/10.1007/978-3-030-41239-5.

[66] Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. SSM, 210:2–21. https://doi.org/10.1016/j.socscimed.2017.12.005.

validated questionnaires assessing depression or PTSD, for example. According to this mechanistic reasoning, these improvements would be caused not only by the drug effect in each patient, x(i), but crucially modulated, positively and/or negatively, by a series of support factors both interacting with the drug, a3b(i), and also the ones independent from the drug, a4z(i). Examples of known support factors include the patient-doctor relationship, therapeutic alliance, patient and therapist expectations, interpersonal relationships, psychotherapeutic interventions, goal-oriented behaviours, language, self-narratives, affective emotional expression, music, supportive or therapeutic touch, family support, socioeconomic status, etc. Interestingly, from this epistemic perspective, as the postulated causal principle includes these support factors as part of the mechanistic reasoning, classifying them as biases would be to reject, *a priori*, the postulated causal principle. In other words, it would be equivalent to rejecting a possible complex and multifaceted mechanism of action, including goal-directed behaviours and social interactions. Furthermore, without explicit logical and epistemic arguments for doing so, this becomes a philosophical bias: "assumptions of a non-empirical nature about topics such as causality, determinism and reductionism when conducting research"[67]. A critical assessment must also take into account that risk of bias is not the same as an actualized bias, and thus risk of bias does not automatically invalidate results. Rather it calls for careful interpretation[68]. In PAT, end of trial blinding assessments likely include noticeable improvements and therefore benign unblinding through efficacy[69]. Invalidating such results would waste potentially effective treatments. Therefore the question for EBPP assessments is not only methodological but also socio-epistemological, and involves assessing if such complex mechanistic reasoning is indeed supported by good quality mechanistic evidence.

In the case of psychedelics, good quality mechanistic evidence includes not only the more proximal receptor binding and intracellular signalling pathways *in vitro* and *in vivo*[70] and downstream neural plasticity[71], but also the reopening of social reward learning periods[72],

---

[67] Andersen, F., Anjum, RL., Rocca, E. (2019) Philosophy of Biology: Philosophical bias is the one bias that science cannot avoid eLife 8:e44929. https://doi.org/10.7554/eLife.44929.

[68] Worrall, J. (2022). Philosophy of Science Meets Medicine (Again): A Clearer-Sighted View of the Virtues of Blinding and of Tests for Blinding in Clinical Trials. In: Gonzalez, W.J. (eds) CURRENT TRENDS IN PHILOSOPHY OF SCIENCE (Synthese Library, vol 462. Springer, Cham.) https://doi.org/10.1007/978-3-031-01315-7_2.

[69] Schenberg E. E. (2021). Who is blind in psychedelic research? Letter to the editor regarding: blinding and expectancy confounds in psychedelic randomized controlled trials. Exp. Rev. Clin. Pharmacol., 14(10):1317–1319. https://doi.org/10.1080/17512433.2021.1951473.

[70] E.g. Hess, E. M., & Gould, T. D. (2023). Possible psychedelic therapeutic mechanism. Science, 379(6633):642–643. https://doi.org/10.1126/science.adg2989; Bakoyiannis, I. A new mechanism of psychedelic action in mice. Nat. Mental Health 1, 450 (2023). https://doi.org/10.1038/s44220-023-00096-y.

[71] E.g. Grieco, S. F., Castrén, E., Knudsen, G. M., Kwan, A. C., Olson, D. E., Zuo, Y., Holmes, T. C., & Xu, X. (2022). Psychedelics and Neural Plasticity: Therapeutic Implications. J. Neurosci., 42(45):8439–8449. https://doi.org/10.1523/JNEUROSCI.1121-22.2022; Calder, A.E., Hasler, G. Towards an understanding of psychedelic-induced neuroplasticity. Neuropsychopharmacol. 48:104–112 (2023). https://doi.org/10.1038/s41386-022-01389-z.

[72] Reardon S. (2023). How psychedelic drugs achieve their potent health benefits. Nature, 618(7966), 654–655. https://doi.org/10.1038/d41586-023-01920-2.

empathy-like[73] and prosocial behaviours[74] in mice; social behaviour in cephalopods[75] and sociality in humans[76]. Research in humans has also evolved a lot in understanding the phenomenology of the psychedelic experience. These drugs trigger a variety of subjective experiences conceptualised as emotional breakthrough[77], connectedness to self, others and world[78], ego-dissolution[79], oceanic-boundlessness[80] and peak affective and cognitive experiences contentiously framed as being mystical or spiritual[81], as well as intensely challenging experiences[82]. Furthermore, some of these subjective experiences can be mapped to neurobiological circuits and neurochemistry of different compounds[83], corroborating their evidential relevance as mechanisms. Finally, some clinical studies showed that the quality of

---

[73] Rein, B., Raymond, K., Boustani, C., Tuy, S., Zhang, J., St Laurent, R., Pomrenze, M. B., Boroon, P., Heifets, B., Smith, M., & Malenka, R. C. (2024). MDMA enhances empathy-like behaviors in mice via 5-HT release in the nucleus accumbens. Science advances, 10(17), eadl6554. https://doi.org/10.1126/sciadv.adl6554.

[74] Bhatt, K.V., Weissman, C.R. The effect of psilocybin on empathy and prosocial behavior: a proposed mechanism for enduring antidepressant effects. Mental Health Res.. 3(7), 2024. https://doi.org/10.1038/s44184-023-00053-8.

[75] Rachel Nuwer. “Rolling Octopuses”. Scientific American Magazine, 319(6):17, 2018. http://doi.org/10.1038/scientificamerican1218-18.

[76] Roseman, L., Preller, K. H., Fotiou, E., & Winkelman, M. J. (2022). Editorial: Psychedelic sociality: Pharmacological and extrapharmacological perspectives. Front. Pharmacol, 13:979764. https://doi.org/10.3389/fphar.2022.979764.

[77] Roseman, L., Haijen, E., Idialu-Ikato, K., Kaelen, M., Watts, R., & Carhart-Harris, R. (2019). Emotional breakthrough and psychedelics: Validation of the Emotional Breakthrough Inventory. J. Psychopharmacol., 33(9):1076–1087. https://doi.org/10.1177/0269881119855974;

[78] Watts, R., Kettner, H., Geerts, D., Gandy, S., Kartner, L., Mertens, L., Timmermann, C., Nour, M. M., Kaelen, M., Nutt, D., Carhart-Harris, R., & Roseman, L. (2022). The Watts Connectedness Scale: a new scale for measuring a sense of connectedness to self, others, and world. Psychopharmacol., 239(11), 3461–3483. https://doi.org/10.1007/s00213-022-06187-5.

[79] Nour, M. M., Evans, L., Nutt, D., & Carhart-Harris, R. L. (2016). Ego-Dissolution and Psychedelics: Validation of the Ego-Dissolution Inventory (EDI). Front. Human Neurosci., 10, 269. https://doi.org/10.3389/fnhum.2016.00269; Letheby, C., & Gerrans, P. (2017). Self unbound: ego dissolution in psychedelic experience. Neurosci. Conscious., 2017(1):nix016. https://doi.org/10.1093/nc/nix016; Stoliker, D., Egan, G. F., Friston, K. J., & Razi, A. (2022). Neural Mechanisms and Psychology of Psychedelic Ego Dissolution. Pharmacol. Rev., 74(4):876–917. https://doi.org/10.1124/pharmrev.121.000508.

[80] Preller, K. H., & Vollenweider, F. X. (2018). Phenomenology, Structure, and Dynamic of Psychedelic States. Cur. Top. Beh. Neurosci., 36:221–256. https://doi.org/10.1007/7854_2016_459.

[81] Pahnke W. N. (1969). Psychedelic drugs and mystical experience. Int. Psyc. Clin., 5(4):149–162; Griffiths, R. R., Richards, W. A., McCann, U., & Jesse, R. (2006). Psilocybin can occasion mystical-type experiences having substantial and sustained personal meaning and spiritual significance. Psychopharmacol., 187(3):268–292. https://doi.org/10.1007/s00213-006-0457-5; Mosurinjohn, S., Roseman, L., & Girn, M. (2023). Psychedelic-induced mystical experiences: An interdisciplinary discussion and critique. Front. Psych., 14:1077311. https://doi.org/10.3389/fpsyt.2023.1077311;

[82] Barrett, F. S., Bradstreet, M. P., Leoutsakos, J. S., Johnson, M. W., & Griffiths, R. R. (2016). The Challenging Experience Questionnaire: Characterization of challenging experiences with psilocybin mushrooms. J. Psychopharmacol., 30(12):1279–1295. https://doi.org/10.1177/0269881116678781;

[83] Zamberlan, F., Sanz, C., Martínez Vivot, R., Pallavicini, C., Erowid, F., Erowid, E., & Tagliazucchi, E. (2018). The Varieties of the Psychedelic Experience: A Preliminary Study of the Association Between the Reported Subjective Effects and the Binding Affinity Profiles of Substituted Phenethylamines and Tryptamines. Front. Int. Neurosc., 12, 54. https://doi.org/10.3389/fnint.2018.00054; Ballentine, G., Friedman, S. F., & Bzdok, D. (2022). Trips and neurotransmitters: Discovering principled patterns across 6850 hallucinogenic experiences. Science adv., 8(11), eabl6989. https://doi.org/10.1126/sciadv.abl6989.

lived experiences[84] and interpersonal relationships[85] correlated with treatment outcomes, corroborating causal claims that these are part of the therapeutic mechanism of action. How much each feature contributes, which are indispensable for different disorders, and the interactions among them are still open for further research, but the appreciation of a complex causal model depends on differing philosophical commitments regarding the experiential domain and lived experiences in psychedelic research[86]. Given EBM's positivist epistemology and practices which "obscures the subjective elements that inescapably enter all forms of human inquiry"[87], it is particularly important to emphasise that "it is not the individuals who have experience, but subjects who are constituted through experience"[88] and that "phenomenological psychopathology – as the basic science of psychiatry – represents an important methodology for advancing evidence-based practices in mental health"[89].

On the other hand, for postulated mechanistic reasoning such as "[t]he effects observed thus far in the best controlled studies of psychedelic treatment must be attributed to the drug itself and not to psychotherapy", where poorly defined "psychological support" is claimed to be for safety reasons only[90], the formalisation of the causal principle might look like:

---

[84] Sloshower, J., Zeifman, R. J., Guss, J., Krause, R., Safi-Aghdam, H., Pathania, S., Pittman, B., & D'Souza, D. C. (2024). Psychological flexibility as a mechanism of change in psilocybin-assisted therapy for major depression: results from an exploratory placebo-controlled trial. Sci. Rep., 14(1):8833. https://doi.org/10.1038/s41598-024-58318-x; Weiss, B., Ginige, I., Shannon, L., Giribaldi, B., Murphy-Beiner, A., Murphy, R., Baker-Jones, M., Martell, J., Nutt, D. J., Carhart-Harris, R. L., & Erritzoe, D. (2024). Personality change in a trial of psilocybin therapy v. escitalopram treatment for depression. Psych. Med, 54(1):178–192. https://doi.org/10.1017/S0033291723001514; Ko, K., Knight, G., Rucker, J. J., & Cleare, A. J. (2022). Psychedelics, Mystical Experience, and Therapeutic Efficacy: A Systematic Review. Front. Psych., 13:917199. https://doi.org/10.3389/fpsyt.2022.917199; Roseman, L., Nutt, D. J., & Carhart-Harris, R. L. (2018). Quality of Acute Psychedelic Experience Predicts Therapeutic Efficacy of Psilocybin for Treatment-Resistant Depression. Front. in Pharmacol., 8, 974. https://doi.org/10.3389/fphar.2017.00974; Majić, T., Schmidt, T. T., & Gallinat, J. (2015). Peak experiences and the afterglow phenomenon: when and how do therapeutic effects of hallucinogens depend on psychedelic experiences? J. Psychopharmacol., 29(3):241–253. https://doi.org/10.1177/0269881114568040.

[85] Murphy, R., Kettner, H., Zeifman, R., Giribaldi, B., Kartner, L., Martell, J., Read, T., Murphy-Beiner, A., Baker-Jones, M., Nutt, D., Erritzoe, D., Watts, R., & Carhart-Harris, R. (2022). Therapeutic Alliance and Rapport Modulate Responses to Psilocybin Assisted Therapy for Depression. Front. Pharmaco., 12, 788155. https://doi.org/10.3389/fphar.2021.788155.

[86] Yaden, D. B., & Griffiths, R. R. (2020). The Subjective Effects of Psychedelics Are Necessary for Their Enduring Therapeutic Effects. ACS Pharmacol. & Transl. Sci., 4(2):568–572. https://doi.org/10.1021/acsptsci.0c00194; Olson D. E. (2020). The Subjective Effects of Psychedelics May Not Be Necessary for Their Enduring Therapeutic Effects. ACS Pharmacol. & Transl. Sci., 4(2), 563–567. https://doi.org/10.1021/acsptsci.0c00192.

[87] Goldenberg M. J. (2006). On evidence and evidence-based medicine: lessons from the philosophy of science. SSM, 62(11): 2621–2632. https://doi.org/10.1016/j.socscimed.2005.11.031;

[88] Joan Scott apud Goldenberg M. J. (2006). On evidence and evidence-based medicine: lessons from the philosophy of science. SSM, 62(11): 2621–2632. https://doi.org/10.1016/j.socscimed.2005.11.031;

[89] Ritunnano R, Papola D, Broome MR, Nelson B. Phenomenology as a resource for translational research in mental health: methodological trends, challenges and new directions. Epidemiol. Psych. Sci. 2023:32:e5. http://doi.org/10.1017/S2045796022000762.

[90] Goodwin, G. M., Malievskaia, E., Fonzo, G. A., & Nemeroff, C. B. (2024). Must Psilocybin Always "Assist Psychotherapy"?. AJP, 181(1): 20–25. https://doi.org/10.1176/appi.ajp.20221043.

CP : *y(i)c = a1 + a2y0(i) + x(i)* (adapted from Cartwright and Hardie, 2012[91])

where support factors related to, and unrelated to, the treatment variable, a3b(i) and a4z(i), respectively, were supposedly controlled by randomisation and blinding and therefore excluded from the causal principle. For this kind of causal claim, however, the epistemic weaknesses of randomisation and blinding are paramount, as in psychedelic trials correct guesses in treatment groups approaches 100%, at least in the 17% of recent studies reporting it[92], and lower rates would be very hard to obtain in experimental groups. In this case, the mechanistic evidence to support this kind of mechanistic reasoning would necessarily include actual evidence that participants were blinded and that other subjective and interpersonal factors were not statistically related to efficacy outcomes. Evidencing such claims requires explicitly reporting and evaluating how much risk of bias was actually present or not.

As in evidential pluralism all evidence must be made explicit and explicitly discussed, for the two types of causal claims related to psychedelics, measuring and reporting blinding assessments, expectations, subjective and phenomenological aspects and therapeutic alliance are necessary. However, in PAT unblinding does not threaten internal validity, given that these factors are here postulated as mechanisms of action. Furthermore, unblinding does not mean that the drug does not contribute to efficacy. It means that the drug is not the only factor related to efficacy. This is why theoretical assumptions such as Muthukumaraswamy's are extremely unlikely. Finally, unblinding and therapy do not seriously threaten external validity in assisted-therapy models, because if the extrapolation aims for other assisted-(psycho)therapy practices and contexts, a similar causal structure can be more reasonably expected. This should be, however, a priority area for funding and more fine-grained research to better assess what are the indispensable contributing factors in different therapeutic approaches, and which derailing factors can undermine extrapolation, for example different socio-economic and cultural variables. On the other hand, for causal claims that efficacy is caused solely by the drug, internal validity is threatened by unblinding, as efficacy may have been strongly influenced by factors other than the drug itself. In addition, external validity is seriously undermined, given that the "drug only" mechanistic reasoning may not be supported by sound mechanistic evidence. Therefore, the causal structure in each RCT may have included many other factors which can't easily be controlled for, making it likely that the causal principle in RCT samples will not be obtained in a variety of target populations with distinct causal structures.

Discussion

As current regulatory practices are still highly reliant on EBM, despite its epistemological shortcomings, the eventual implementation of psychedelics as therapeutics requires serious

[91] Cartwright, N. and Hardie, J. EVIDENCE-BASED POLICY: A PRACTICAL GUIDE TO DOING IT BETTER (New York, 2012), https://doi.org/10.1093/acprof:osobl/9780199841608.001.0001;
[92] Nayak, S. M., Bradley, M. K., Kleykamp, B. A., Strain, E. C., Dworkin, R. H., & Johnson, M. W. (2023). Control Conditions in Randomized Trials of Psychedelics: An ACTTION Systematic Review. J. Clin. Psych., 84(3):22r14518. https://doi.org/10.4088/JCP.22r14518.

consideration of limitations in current evidentiary norms[93]. If these considerations continue to be overlooked during assessments of psychedelic therapies, there are increased risks of rejection of effective treatments, as well as scaling up impoverished and riskier practices[94]. This can ultimately lead to a medical reversal: approved treatments later abandoned by being perceived as ineffective and/or harmful[95]. In the case of psychedelic therapy, however, an eventual medical reversal would not be caused by lack of rigour in RCTs, but blindness to the extrapolation fallacy obscuring the study of the myriad support factors and complex causal structures of therapies during non-ordinary states of consciousness[96].

Acknowledgements

I want to express deep gratitude to the Petrie-Flom Center and participants at the Annual Conference in June 2024 for fruitful discussions on an early version of this manuscript. And special thanks to Franklin King and Michela Massimi for their support to my philosophical inquiries.

[93] Hendy, K. (2018). Placebo Problems: Boundary Work in the Psychedelic Science Renaissance. In: Labate, B., Cavnar, C. (eds) PLANT MEDICINES, HEALING AND PSYCHEDELIC SCIENCE (Springer, Cham). https://doi.org/10.1007/978-3-319-76720-8_9. McGuire, A. L., Lynch, H. F., Grossman, L. A., & Cohen, I. G. (2023). Pressing regulatory challenges for psychedelic medicine. Science (New York, N.Y.), 380(6643), 347–350. https://doi.org/10.1126/science.adg1324.
[94] Schenberg, E. E., King, F., 4th, da Fonseca, J. E., & Roseman, L. (2024). Is Poorly Assisted Psilocybin Treatment an Increasing Risk?. AJP, 181(1):75–76. https://doi.org/10.1176/appi.ajp.20230664.
[95] Cifu, AS. and VK. Prasad. ENDING MEDICAL REVERSAL: Improving Outcomes, Saving Lives (Johns Hopkins University Press, 2015). https://doi.org/10.1353/book.49286.
[96] Schenberg E. E. (2018). Psychedelic-Assisted Psychotherapy: A Paradigm Shift in Psychiatric Research and Development. Front. Pharmacol., 9:733. https://doi.org/10.3389/fphar.2018.00733.