Journal of Anomalistics
Zeitschrift für Anomalistik
Vol. 26 (2026), Issue 1

ISSN 2747-9315

www.anomalistik.de www.igpp.de

26 • 1

Journal of Anomalistics
Zeitschrift für Anomalistik

Vol. 26, Issue 1

2026

Journal of Anomalistics • Zeitschrift für Anomalistik • 2026

Gesellschaft für Anomalistik

Institut für Grenzgebiete der Psychologie und Psychohygiene

Impressum

ISSN 2747-9315

www.anomalistik.de
www.igpp.de

Satz / Typesetting: Kirsten Krebber / Gerhard Mayer
Herausgeber / Publishers: Gesellschaft für Anomalistik
Institut für Grenzgebiete der Psychologie und Psychohygiene
Erscheinungsort / Place of Publication: Freiburg im Breisgau

Profile of the Journal / Profil der Zeitschrift

The *Journal of Anomalistics* sees itself as a scientific forum for promoting controversial discourse on scientific anomalies, extraordinary human experiences, and so-called parasciences. Empirical research reports, general papers on methodological, conceptual, philosophical, or history of science aspects, review articles, commentaries and discussion papers, and book reviews are published. Guiding research questions on scientific anomalies, extraordinary human experiences, as well as parasciences are those on truth content and explanatory models, on the psychosocial backgrounds of the associated belief systems, and on the social frameworks of anomaly-provoked advances in knowledge in science. Methodological pluralism, competing scientific theoretical approaches, and interdisciplinary approaches are encouraged.

Die *Zeitschrift für Anomalistik* versteht sich als ein wissenschaftliches Forum zur Förderung eines kontroversen Diskurses über wissenschaftliche Anomalien, außergewöhnliche menschliche Erfahrungen und sog. Parawissenschaften. Veröffentlicht werden empirische Forschungsberichte, allgemeine Abhandlungen zu methodischen, konzeptuellen, philosophischen oder wissenschaftshistorischen Aspekten, Review-Artikel, Kommentare und Diskussionsbeiträge, sowie Buchrezensionen. Leitende Forschungsfragen zu wissenschaftlichen Anomalien, außergewöhnlichen menschlichen Erfahrungen sowie Parawissenschaften sind die nach Wahrheitsgehalt und Erklärungsmodellen, nach den psychosozialen Hintergründen der damit verbundenen Überzeugungssysteme, sowie nach den sozialen Rahmenbedingungen von durch Anomalien provozierten Erkenntnisfortschritt in der Wissenschaft. Methodenpluralismus, konkurrierende wissenschaftstheoretische Ansätze, sowie interdisziplinäre Zugänge sind erwünscht.

***

Journal of Anomalistics • Zeitschrift für Anomalistik

Volume 26 (2026), Number 1

Contents / Inhalt

Continued discussions on previous contributions

Fortgesetzte Diskussionen zu früheren Beiträgen

Obituaries / Nachrufe

Book Reviews / Rezensionen

Abstracts-Dienst / Literaturspiegel

Journal of Anomalistics
Zeitschrift für Anomalistik

www.anomalistik.de www.igpp.de

Editorial

"Effort Structure" and Superhumanism

Gerhard Mayer
mayer@anomalistik.de

For several years, from 1987 to 1994, I worked as a sculptor for the artist Helmut Lutz, who primarily created sculptures and installations for churches and public spaces. His detailed, figurative, and almost Baroque-style works did not necessarily align with the zeitgeist. Helmut Lutz drew inspiration from figurative sculpture and the symbolism of High Gothic art. He called one of his concepts, which he frequently sought to convey to his students, "effort structure" ["Mühestruktur"]. He did not place primary emphasis on the sudden flash of inspiration – the conceptual idea – where the actual physical realization and shaping of the object were viewed as secondary. At least as significant was the effort, in the sense of the labor involved, that flows into the craftsmanship of the art object. The time spent on the object during its creation should be palpable in the design; it should constitute part of its impact, its aura. This is certainly an idealistic approach, a form of devotion to the work that seeks not immediate, rapid success, but rather a sustained engagement with form and material.

After all these years, Lutz' neologism "effort structure" resurfaced in my memory quite unexpectedly as I was finalizing this issue of the *Journal of Anomalistics*. The process leading up to the final compilation of the contributions was marked by several last-minute decisions that resulted in an unplanned thematic focus for the issue: the presentation and discussion of the work conducted by the research group led by Markus Maier, a professor for psychology at Ludwig-Maximilians-University Munich.

The approach adopted by Markus Maier, Moritz Dechamps, Anastasia Vogel, and Johannes Storch takes Generalized Quantum Theory and its implications as the research-guiding theoretical paradigm. In this respect, it breaks new ground in parapsychology, as the anomalies hypothesized and documented do (presumably) not constitute classical psi phenomena. For this reason, terms that often act as buzzwords – such as extrasensory perception (ESP) or psychokinesis (PK) – do not appear in their texts. Perhaps this will result – or at least, one would hope so – in a less reflexive defensive reaction from the scientific mainstream.

Maier and colleagues submitted the first paper well in advance of publication deadlines, and the dedicated reviewers worked quickly as well. The depth and substantive quality of the reviews led to the decision to initiate an open commentary process on the article. In the meantime, Maier et al. had completed another paper for submission – a study that built directly on the first. Since the upcoming December issue of journal is scheduled to become a special issue on "Parapsychology under Socialism," the editorial board decided to take the unconventional step of including the second paper in this issue as well. The decision to bundle two articles and permit a discussion without word-count limits for the comments and the authors' response runs counter to the standardized format of many academic journals and poses a challenge for the shortened attention spans of our time. For readers, this may require a considerable effort and some diligence to work through the texts, but it is well worth the effort for those interested in new developments in experimental parapsychology.

When viewing all the texts – the two main articles, the three commentaries, and the authors' response, comprising over 130 printed pages in total – as a whole, an "effort structure" becomes apparent. I observed this throughout the entire process among all those involved, in their constructive debates and in their dedication to research work, that does not seek easy shortcuts yet proves to be extremely effective.

In the studies by Wahbeh et al. and Laythe & Houran presented in this issue, too, one can sense that no effort was spared to advance their research. For those familiar with these research groups, this is evident anyway.

It would be a mistake to equate the incorporation of an "effort structure" into one's work with a rejection of technical aids. As we all know, artificial intelligence (AI) has become part of our everyday lives; accordingly, it is also part of the everyday routine and toolkit of scientists. However, it would be an error to believe that AI allows one to arrive at substantial scientific insights conveniently and effortlessly. The decisive factor is the time invested by the human researcher. A thorough analysis of problems and research questions, comprehensive research into the current state of the field, careful definition of research variables, reflection on the results

and their limitations, and constructive exchange with colleagues – all of these are fundamental elements of sound scientific work that should be delegated to artificial assistants only to a limited extent. That is why one should not really speak of the "implementation of an effort structure," as I did above, because such effort is an intrinsic part of scientific activity.

Just as the term "Mühestruktur" recently surfaced from the depths of my memory, I happened to come across a thought-provoking and, it seems to me, thematically relevant article in the web magazine *The Point*. The text, written by scholar of religion Nicholas Low, is titled "Superhuman Fantasies: Nietzsche versus Techno-Optimists" (Low, 2026). The author describes the visions of optimistic transhumanists, according to whom humanity will transcend its present nature through a variety of future technologies.[1] This goal is to be pursued at an accelerated pace via unbridled techno-capitalism, in line with the ideology known as accelerationism. In Marc Andreessen's *Techno-Optimist Manifesto*, the section titled "Enemies" states: "Our enemy is Friedrich Nietzsche's Last Man" (Andreessen, 2023). The "Last Man" from Nietzsche's *Thus Spoke Zarathustra* is meant to be superseded by the Übermensch (superhuman). Nicholas Low points out the misinterpretation of Nietzsche by such tech optimists, who reverse the meaning of the term "Übermensch" as Nietzsche understood it.

While the "tech CEOs" of Silicon Valley state the abolition of human suffering through the creation of the superhuman as their goal,

> Nietzsche dreamed of a superhuman for whom hardship, including the hardship of finitude, would be no objection to life. The Übermensch would overcome the fear and hatred of suffering, not seek to leave suffering behind. What we see in Silicon Valley superhumanism is not the way out of nihilism but its most dangerous instantiation.

And further:

> If Nietzsche was correct that we need the superhuman to overcome nihilism, then these technological superhumanists are effectively prescribing for the disease of nihilism the amplification of nihilistic values. Andreessen's "technological supermen" would not be Übermensch but the "last men" par excellence. (Low, 2026)

To return to the mundane sphere of everyday practical activities of human beings: using an airport shuttle service when one could easily walk the distance, having groceries and ready-made meals delivered to one's door, or letting a smart home and 24-hour delivery services handle the need for planning and organization – all this sounds like luxury, a comfortable and enviable

1 I do not wish to address the many problematic ethical implications associated with this agenda here.

situation in a postmodern world. Yet, the resulting lack of physical exertion produces unpleasant side effects that are all too often ignored. A lack of physical activity, to cite a well-known example, quite rapidly leads to muscle atrophy. Archaeological findings show that in premodern times, children from privileged classes exhibited skeletal damage indicative of vitamin deficiency despite a plentiful diet (Snoddy et al., 2024). Exposing oneself to sunlight was something for the lower classes; eating white bread was better and more befitting one's station than eating brown bread and groats, and cake was even better …

In a recent interview, computer scientist Cal Newport had this to say about the dangers of a poor "cognitive diet" or complacency that implies avoiding effort structures:

> Academic institutions need to demonstrate that the life of the mind is hard and worth it. We need to think about cognitive fitness the way we think about physical fitness. There should be a simple rule for being a thinker in an age of AI: Don't let AI write anything for you. Writing is to cognitive health what steps are to physical health. Write that email from scratch. Write that memo with the bullet points from scratch. Don't flee that strain. You need it as much as you need those 10,000 steps a day. (Goldstein, 2026)

Against the backdrop of these cultural-critical reflections, the concept of the "effort structure" can be viewed in a slightly different light – as an antidote to an overheated academic system dominated by the "publish-or-perish" principle and the highly problematic developments resulting from it. This by no means precludes effectiveness, as the studies presented in this issue of the *Journal of Anomalistics* demonstrate.

References

Andreessen, M. (2023, October 16). The Techno-Optimist Manifesto. Andreessen Horowitz. https://a16z.com/the-techno-optimist-manifesto/

Gauquelin, M., & Brown, F. A. (1969). The cosmic clocks from astrology to a modern science (First British Commonwealth edition). Owen.

Goldstein, E. (2026). Is AI Making Us Stupid? Cal Newport Is Worried. *The Chronicle of Higher Education, 72*(15). https://www.chronicle.com/article/is-ai-making-us-stupid-cal-newport-is-worried

Low, N. (2026, May 27). Superhuman Fantasies. *The Point Magazine*. https://thepointmag.com/examined-life/superhuman-fantasies/

Snoddy, A. M. E., Shaw, H., Newman, S., Miszkiewicz, J. J., Stewart, N. A., Jakob, T., Buckley, H., Caffell, A., & Gowland, R. (2024). Vitamin D status in post-medieval Northern England: Insights from dental histology and enamel peptide analysis at Coach Lane, North Shields (AD 1711–1857). *PLOS ONE, 19*(1), e0296203. https://doi.org/10.1371/journal.pone.0296203

Editorial: „Mühestruktur“ und Superhumanismus

Ich arbeitete von 1987 bis 1994 als Bildhauer für den Künstler Helmut Lutz, der hauptsächlich Skulpturen und Raumgestaltungen für Kirchen und den öffentlichen Raum realisierte. Seine detailreichen, figurativen und fast barock anmutenden Werke entsprachen nicht unbedingt dem Zeitgeist. Helmut Lutz war von Bildwerken und von dem Symbolismus der Hochgotik inspiriert. Eines seiner Konzepte, welches er nicht müde wurde, seinen Schülern verständlich zu machen, nannte er „Mühestruktur“. Dem zündend-neuen Gedankenblitz, der Konzeptidee, bei der die konkrete Realisierung und Gestaltung in einem physikalischen Objekt als sekundär angesehen wurde, wurde nicht das Hauptgewicht beigemessen. Mindestens ebenso bedeutsam war die Mühe im Sinne des Arbeitsaufwands, die in die handwerkliche Ausarbeitung des Kunstobjekts einfließt. Die Zeit, die mit dem Objekt bei dessen Herstellung verbracht wird, soll in der Gestaltung spürbar werden, soll einen Teil seiner Wirkung oder Aura ausmachen. Das ist sicher ein idealistischer Ansatz, eine Form der Hingabe an ein Werk, wo nicht nach unmittelbarem und schnellem Erfolg gestrebt, sondern die beständige Auseinandersetzung mit Form und Material gesucht wird.

Eine Wortschöpfung, ein (Kunst-)Begriff, der in einer biografisch weit zurückliegenden Phase in meinem Leben bedeutsam war, tauchte völlig unerwartet bei der Fertigstellung dieser Ausgabe der *Zeitschrift für Anomalistik* in meiner Erinnerung auf. Der Prozess bis zur endgültigen Zusammenstellung der Beiträge war gekennzeichnet von einigen kurzfristigen Entscheidungen, die zu einem ungeplanten thematischen Schwerpunkt der Ausgabe führten, nämlich die Präsentation und Diskussion der Arbeit der Forschungsgruppe um den Professor für Psychologie Markus Maier von der Ludwig-Maximilians-Universität München.

Der Ansatz von Markus Maier, Moritz Dechamps, Anastasia Vogel und Johannes Storch wählt die Generalized Quantum Theory mit deren Implikationen zum forschungsleitenden theoretischen Paradigma und geht insofern neue Wege in der Parapsychologie, als die hypothesierten und nachgewiesenen Anomalien (vermutlich) keine klassischen Psi-Phänomene darstellen. Deswegen tauchen die oft als Reizworte wirkenden Begriffe wie außersinnliche Wahrnehmung (ASW) oder Psychokinese (PK) in ihren Texten nicht auf. Vielleicht hat dies – zumindest wäre es zu hoffen – ein geringeres Maß an reflexhaften Abwehrreaktionen im wissenschaftlichen Mainstream zur Folge.

Den ersten Beitrag hatten Maier und Kolleg:innen frühzeitig eingereicht, was die Deadlines für die Publikationen betraf, und auch die engagierten Gutachter arbeiteten schnell. Die Ausführlichkeit und substanzielle Qualität der Gutachten führten zur Überlegung, eine offene Kommentierung des Artikels einzuleiten. Inzwischen hatten Maier et al. ein weiteres Paper zur

Einreichung fertiggestellt – eine Studie, die direkt auf der ersten eingereichten Arbeit aufbaute. Da die nachfolgende Dezemberausgabe der *ZfA* als ein Themenheft zu „Parapsychology under Socialism“ geplant ist, entschied sich die Redaktion zu dem unkonventionellen Schritt, auch das zweite Paper in die vorliegende Ausgabe aufzunehmen. Die Entscheidung, zwei Artikel zu bündeln und eine Diskussion ohne Textlängenbegrenzungen für die Kommentare und die Autorenantwort zuzulassen, steht dem Einheitsformat vieler wissenschaftlicher Fachzeitschriften entgegen und stellt zeitgeistbedingte verminderte Aufmerksamkeitsspannen vor Herausforderungen. Für die Leserschaft mag es einen anspruchsvollen Kraftakt bedeuten und ihr einige Mühen abverlangen, die Texte durchzuarbeiten, doch es lohnt sich, wenn man sich für neue Entwicklungen in der experimentellen Parapsychologie interessiert.

Betrachtet man alle Texte – die beiden Hauptartikel, die drei Kommentare sowie die Autorenantwort mit zusammengenommen über 130 Druckseiten – in ihrer Gesamtheit, dann wird daran eine „Mühestruktur“ sichtbar. Ich konnte sie feststellen in dem ganzen Prozess bei allen Beteiligten in den konstruktiven Auseinandersetzungen und in ihrer Hingabe an die Forschungsarbeit, die nicht nach bequemen Abkürzungen schielt und sich trotzdem extrem effektiv zeigt.

Auch bei den in dieser Ausgabe präsentierten Studien von Wahbeh et al. und Laythe & Houran spürt man, dass keine Mühen gescheut wurden, um ihre Forschung voranzutreiben. Für diejenigen, die diese Forschergruppen kennen, ist dies ohnehin klar.

Es wäre ein Missverständnis, wenn man die Implementierung einer „Mühestruktur“ in die Arbeit mit einer Ablehnung des Verwendens technischer Hilfsmittel gleichsetzen würde. Künstliche Intelligenz (KI) ist, wie wir alle wissen, zu einem Teil unseres Alltags geworden. Dementsprechend gehört sie auch zum Alltag von Wissenschaftler:innen und zu ihrem Werkzeugkasten. Doch wäre es ein Irrtum, zu denken, man könnte mittels KI bequem und mühelos zu substanzieller wissenschaftlicher Erkenntnis gelangen. Der entscheidende Faktor ist die investierte Zeit auf menschlicher Seite. Das gründliche Durchdenken von Problem- und Fragestellungen, die umfassende Recherche zum aktuellen Stand des Fachgebiets, die sorgfältige Definition von Forschungsvariablen, die Reflexion der Ergebnisse und ihrer Grenzen sowie der konstruktive Austausch mit Kollegen – all dies sind grundlegende Elemente guter wissenschaftlicher Arbeit, die nur in begrenztem Umfang an künstliche Helfer delegiert werden sollten. Deswegen sollte man eigentlich gar nicht von der „Implementierung einer Mühestruktur“ sprechen, wie ich es oben getan hatte, denn sie gehört fundamental zur wissenschaftlichen Tätigkeit.

Während der Begriff „Mühestruktur“ aus den Tiefen meiner Erinnerung auftauchte, bin ich zufällig auf einen anregenden und, wie mir scheint, thematisch passenden Artikel im Web-Magazine *The Point* gestoßen. Der Text des Religionswissenschaftlers Nicholas Low trägt den

Titel „Superhuman Fantasies. Nietzsche versus Techno-Optimists“ (Low, 2026). Der Autor beschreibt die Vorstellungen von optimistischen Transhumanisten, nach denen der Mensch mittels vielfältiger Zukunftstechnologien in seiner gegenwärtigen Natur überwunden werden wird.[2] Dieses Ziel soll mittels eines ungebremsten Technokapitalismus in beschleunigter Weise verfolgt werden, wie es der Ideologie des sogenannten Akzelerationismus entspricht. Im *Techno-Optimist Manifesto* von Marc Andreessen heißt es im Abschnitt „Enemies“: „Our enemy is Friedrich Nietzsche's Last Man“ (Andreessen, 2023). Der „letzte Mensch“ aus Nietzsches *Also sprach Zarathustra* soll durch den „Übermenschen“ überwunden werden. Nicholas Low weist auf die Missinterpretation Nietzsches durch solche Tech-Optimisten hin, die die Bedeutung des Begriffs „Übermensch“ in Nietzsches Verständnis ins Gegenteil verkehrt.

Während die „tech CEOs“ des Silicon Valley die Abschaffung menschlichen Leidens durch die Schaffung des „Supermenschen“ als Ziel angeben, träumte

> Nietzsche (…) von einem Übermenschen, für den die Mühsal – einschließlich der Mühsal der Endlichkeit – kein Einwand gegen das Leben wäre. Der Übermensch würde die Angst vor dem Leiden und den Hass darauf überwinden, anstatt zu versuchen, das Leiden hinter sich zu lassen. Was wir im Superhumanismus des Silicon Valley sehen, ist nicht der Ausweg aus dem Nihilismus, sondern dessen gefährlichste Ausprägung.

Und weiter:

> Wenn Nietzsche Recht hatte, dass wir den Übermenschen brauchen, um den Nihilismus zu überwinden, dann verschreiben diese technologischen Superhumanisten gegen die Krankheit des Nihilismus faktisch eine Verstärkung nihilistischer Werte. Andreessens „technologische Supermenschen“ wären keine „Übermenschen“, sondern die „letzten Menschen“ par excellence. (Low, 2026)[3]

Um zur profanen Sphäre der alltäglichen praktischen Tätigkeiten des Menschen zurückzukehren: Wenn man im Flughafen den Fahrdienst in Anspruch nimmt, obwohl man die Strecken gut zu Fuß bewältigen könnte, wenn man sich Einkäufe und fertige Gerichte nur noch nach Hause liefern lässt, wenn einem das Smart-Haus und die 24h-Verfügbarkeit von Lieferservices Planungs- und Organisationsnotwendigkeit abnehmen, dann klingt das nach Luxus und einer bequemen, schönen Situation in einer postmodernen Welt. Doch der damit verbundene Mangel an Mühsal erzeugt unangenehme Begleiterscheinungen, die allzu häufig ignoriert werden.

2 Ich möchte hier nicht auf die zahlreichen problematischen ethischen Implikationen eingehen, die mit dieser Agenda verbunden sind.

3 Alle Übersetzungen durch G. M., mit Unterstützung durch DeepL und Google Translator.

Bewegungsmangel führt, um ein allbekanntes Beispiel zu nennen, recht schnell zu Muskelschwund. Archäologische Funde zeigen, dass in vormodernen Zeiten Kinder in privilegierten Schichten Skelettschädigungen aufwiesen, die trotz üppiger Ernährung auf Vitaminmangel hinweisen (Snoddy et al., 2024). Sich dem Sonnenlicht auszusetzen, war etwas für das niedere Volk; sich von Weißbrot zu ernähren war besser und standesgemäßer als von Graubrot und Grütze, und Kuchen war noch besser …

In einem Interview äußerte sich der Computerwissenschaftler Cal Newport jüngst folgendermaßen zu den Gefahren einer falschen „kognitiven Diät“ oder einer Bequemlichkeit, die Anstrengungen aus dem Weg geht:

> Wissenschaftliche Einrichtungen müssen zeigen, dass geistige Arbeit zwar anspruchsvoll, aber die Mühe wert ist. Wir sollten geistige Fitness genauso betrachten wie körperliche Fitness. Für das Denken im Zeitalter der KI sollte eine einfache Regel gelten: Lass dir von der KI nichts schreiben. Schreiben ist für die geistige Gesundheit das, was Schritte für die körperliche Gesundheit sind. Verfasse die E-Mail von Grund auf selbst. Erstelle das Memo mit den Stichpunkten eigenständig. Weiche dieser Anstrengung nicht aus. Du brauchst sie ebenso sehr wie die 10.000 Schritte am Tag. (Goldstein, 2026)

Vor dem Hintergrund dieser kulturkritischen Überlegungen kann man das Konzept der „Mühestruktur“ noch in einem etwas anderen Licht sehen – als ein Antidot für einen heiß gelaufenen Wissenschaftsbetrieb mit dem ihn dominierenden „Publish-or-perish“-Prinzip und den daraus resultierenden höchst problematischen Entwicklungen. Dies schließt Effektivität keineswegs aus, wie die in dieser Ausgabe der *Zeitschrift für Anomalistik* vorgestellten Studien zeigen.

(Die Literaturliste befindet sich am Ende der englischen Version des Editorials auf Seite 9.)

Journal of Anomalistics
Zeitschrift für Anomalistik

www.anomalistik.de www.igpp.de

Non-Classical Correlation Between Subjective and Objective Color Observations

Change of Effect as a Function of its Empirical Documentation

Markus A. Maier,[a,*] Anastasia Vogel,[a] Johannes Storch,[a] Moritz C. Dechamps [a]

(a) Department of Psychology
Ludwig-Maximilians-Universität, München, Germany

[*] Corresponding author:
Markus Maier
markus.maier@psy.lmu.de

Abstract – In recent research testing the non-commutability conjecture of the Generalized Quantum Theory, Maier and Dechamps (2025) reported macroscopic complementary relations empirically documented as non-classical correlations between subjective assessment of likability of colors and variations of objective documentations of these colors. Specifically, they found higher likability mean scores in a condition where objective color parameters were not erased and stored in a result file compared to a condition in which these parameters were erased and inaccessible by experimenters. This effect was robust across four studies. Each study's design in this series of experiments was intentionally confounded with a "color bias" variable. Only limited sets of colors after random creation were specifically assigned to one of the two conditions, so that the effect of the erasure manipulation could alternatively be explained by the specific features of the colors assigned to conditions (the confound was later statistically ruled out by permutations). The presence of a confound was introduced to avoid the impact of the NT axiom, which forbids a strict causal testing approach. Rather, according to the NT axiom when macroscopic complementary relations are tested in a strictly causal way, initial effect documentations should be followed by a decline within and across studies. In the study presented here, the studies of Maier and Dechamps (2025) were conceptually replicated by a strict, i.e. unconfounded, causal testing

strategy, to explicitly test the proposed impact of the NT axiom. In two experimental conditions (within-subjects) objective color parameters were either stored or not stored. Assignment of colors to conditions was randomized on the participant level this time ensuring an unconfounded design. The prediction derived from the NT axiom was that under such circumstances the erasure-dependent likability effect should initially be found and later drastically decline. The results supported this prediction: A higher likability mean score was initially found in the non-erasure compared to the erasure condition with strong Bayesian evidence (BF_{10} = 39.77), replicating the previous findings but now followed by a decline within the study. These findings are in line with the conjectures of the GQT. Specifically, they indicate that subjective assessments and objective measurements of colors are non-commutable and that the data follow the predictions of the NT axiom exhibiting an "effect and decline" data pattern when strictly causally tested. An additional permutation analysis showed that the probability to find such an effect and decline data pattern by chance was very low ($p < .03$) indicating that it could hardly be interpreted as "false positive" finding. The impact of this and the previous results on the validity of the GQT in describing psychophysical relations as non-classical correlations between spatio-temporally separated subjective assessments and objective measurements of colors is discussed.

Keywords: Generalized Quantum Theory, non-transmission axiom, NT axiom, macroscopic complementarity, macroscopic non-local entanglement correlation, psychophysical interaction, subjective-objective duality

Introduction

The relationship between subjective experiences known as "qualia" (Lewis, 1929) and the objective physical world is often summarized under the umbrella term "psychophysical problem." This topic has been highly controversial for centuries and continues to be so today (see Ruffing, 2021). Some (e.g., Chalmers, 1995; Shariff et al., 2008) argue that conscious experience is incompatible with a physicalist world view (Stoljar, 2024) or that there is at least a fundamental explanatory gap between subjective experiences and objective physical processes accompanying them (Chalmers, 1995, 1996; Levine, 1983; Nagel, 1974). Among the theories that try to close this gap, the Generalized Quantum Theory (GQT) of Walach and Römer (2000, 2011) and Atmanspacher et al. (2002; see also Fach, 2011; Filk & Römer, 2011; Hinterberger & von Stillfried, 2013; Römer, 2023a, 2023b; Lucadou et al., 2007; Walach & von Stillfried, 2011) provides a provocative and unusual solution to the psychophysical problem. In short, instead of postulating causal relationships between subjective and objective reality as is proposed in substance dualism (Descartes, 1641) or instead of reducing consciousness to its physical substrate, the GQT proposes a macroscopic complementary relationship between the subjective and the

objective realms. The nature of this relationship is acausal and can take on the form of macroscopic non-local entanglement correlations (Römer, 2023b). Macroscopic complementarities of this kind constitute systems characterized by mutual co-existence of subjective experiential and objective physical states. Both states are defined by specific observational procedures. Due to their interrelation, measurements of one state determine measurement outcomes of the other state thus predicting non-classical correlations between both measurements in composite systems, analogous to complementarities known in quantum entangled systems (e. g., EPR pairs) that produce non-local correlation structures (Aspect et al., 1982), or single-system complementarities such as the Heisenberg uncertainty relation (Heisenberg, 1927) or the wave-particle duality (Bohr, 1928, 1948, 1949). The term non-classical correlation is used here to distinguish interrelations of macroscopic variables derived from psychophysical subsystems from those that can be explained by purely locally connected variables and their measurements within classically physical systems.

Recently, Maier and Dechamps (2025) provided preliminary evidence for macroscopic complementary relations. Specifically, they reported non-classical correlations between subjective evaluations of color stimuli on the participant level and variations of their objective measurements on the experimenter level, which cannot be explained within the framework of a purely classically physical system. The storage or deletion of objective color parameters, of which participants were unaware, positively correlated with participants' subjective likability ratings of the colors presented. To address the acausal nature of the phenomenon under study, Maier and Dechamps (2025) developed a testing strategy that reduced the causal explanatory power of their experimental design to enable finding robust effects. Acausal macroscopic complementarities between spatiotemporally separated subsystems can take on the characteristics of entanglement correlations between assessments of subjective experiences and objective measurement outcomes. Macroscopically entangled measurement states however cannot be documented robustly in a causal manner according to the non-transmission (NT) axiom of GQT (since this would allow causal signaling transcending the limits of the velocity of light leading to time paradoxes; see Lucadou et al., 2007; Römer, 2023b). Rather, causal experimental tests of these phenomena will lead to a data pattern reflecting an initial effect followed by a decline of the effect with continued data collection (Römer, 2023b). That is, causal tests destroy acausal effects over time, eventually producing a data pattern that closely resembles a false-positive effect. The goal of the study presented here was to test the NT conjecture of the GQT by testing macroscopic complementary relations, which establish non-classical correlations between subjective color experiences and objective color measurements, with a maximum causality test approach. It is thus a conceptual replication of the Maier & Dechamps (2025) studies, but utilizing a maximized instead of reduced causality testing strategy. For such a design the authors

predicted an effect and decline pattern (Maier & Dechamps, 2025, p. 21), but did not provide empirical evidence for their claim. In the following, we will briefly describe the theoretical background of the research, the testing strategy and the central results presented by Maier and Dechamps (2025) before addressing the present study with its direct test of the NT axiom.

Originally, the GQT was designed to describe complementary phenomena beyond the primary domain of physics, such as psychophysical interactions in psychology (Römer, 2023b; Walach & Römer, 2011). Central conjectures of GQT are based on standard quantum theory while abandoning certain conjectures and restrictions that limited its applications to the physical domain only. It is a theory that describes observations derived from different but meaningfully connected subsystems and their interrelations. These observations go beyond purely physical states and might include psychophysical relations between subjective experiences and objective physical measurements of perceptual stimuli. Central to the GQT are the concepts of macroscopic complementarity, entanglement, and observation-dependency of reality descriptions. The term "macroscopic" refers to the fact that these concepts are not limited by the Planck constant, yet they possess the same characteristics as the concepts described in standard quantum theory (Römer, 2023b). According to the GQT, when a color stimulus is presented, subjective experiences are not causally evoked by the physical presence of an object. Rather, the subjective and objective elements of a psychophysical system constitute subsystems that are considered complementary to each other. Each subsystem is established by a distinct form of observation. The subjective experience is derived from a first-person perspective and the objective state is measured physically in a third person-perspective. In case both subsystems are meaningfully connected – e.g., through the same stimulus used to obtain the respective measurements – the observations made on both subsystems can be incommensurable or incompatible due to their complementary nature (Römer, 2023b). This implies that two observables, each derived from the respective subsystem, do not simultaneously provide specific eigenvalues. Rather, the act of measurement of one subsystem changes the state of the whole system including the eigenvalues of the other subsystem. If both complementary observations are derived from spatiotemporally separated subsystems, their measurements are supposed to interact non-locally forming entanglement correlations. The nature of such a relationship between subjective assessments and objective measurements of color stimuli can thus be described as acausal interrelations or non-classical correlations since none can be described as being exclusively caused by its complementary counterpart although both co-relate to each other (Römer, 2023b).

If a psychophysical relation of this kind is subjected to a causal manipulation, this macroscopic complementarity, in this context empirically tested as a non-classical correlation between measurements, – if robust and/or replicable – could be used to transfer signals that

would violate the restrictions posed by special relativity (Lucadou et al., 2007; Römer, 2023b). Thus, the authors of the GQT introduced the NT axiom to avoid this problematic issue by prohibiting a robust local causal use of macroscopic non-local entanglement correlations (Römer, 2023b). As a consequence, psychophysical interactions when subjected to a causal test should display an effect and decline data pattern in the data as described in the Model of Pragmatic Information (MPI; Lucadou et al., 2007). The MPI essentially states that macroscopic non-local entanglement correlations cannot be robustly documented. Although effects should initially appear when causal tests are applied, their robustness weakens with the search for additional evidence and confirmation. Thus, such non-classical entanglement correlations can be detected initially but not robustly which makes them phenomenologically identical to a false-positive data pattern that occurs by chance. Since such data cannot be used for systematic signal transfer they do not pose a challenge to the NT axiom.

Maier and Dechamps (2025) tested the conjectures of the GQT within a series of four studies. Subjective experiences of likability of colors presented on a computer screen were assessed. In addition, the objective color parameters hue and lightness describing these color stimuli on a physical level were either stored in a result file or omitted from the result file and permanently erased after completion of each study (erasure manipulation). This erasure manipulation was performed at the experimenter level after data collection, and participants were unaware of it. This manipulation constituted the independent variable (IV). If subjective and objective assessments are acausally related by macroscopic complementarity, as proposed by the GQT, they should be incommensurable. This means, the measurement of one subsystem's parameters will mutually affect the measurements of the other subsystem. In this case, the erasure manipulation of objective color descriptions should non-classically correlate with observations of their subjective counterparts. Testing the effect of objective color parameter erasure on subjective likability experimentally would constitute a causal test of a proposedly underlying acausal correlation. If a strict causality test would have been performed the effect could not have been documented robustly across studies or even within one study according to the NT axiom. Thus, the authors introduced a confound: In each single study limited subsets of colors were randomly chosen beforehand and assigned to the erasure conditions in a fixed way. The erasure effect could then be alternatively explained by the color features specifying each color subset uniquely assigned to the erasure or non-erasure sub-conditions, respectively (in other words: by a lucky assignment of more likeable colors to one condition). This "biased colors" confound reduced the causal explanatory power of the erasure manipulation by reducing its internal validity. This strategy was supposed to stabilize the erasure effect as non-classical correlation by avoiding the impact of the NT axiom. In all four studies strong Bayesian evidence for higher likability mean scores in the non-erasure compared to the erasure condition was documented, although in each

study different subsets of colors were used. Since each study's data interpretation was confined by the confound no decisive evidence for the existence of a non-classical correlation between subjective color assessments and objective color measurements on the single study-level could be confirmed. Thus, in a final analysis step all data obtained from the four studies were subjected to a permutation analysis which randomly re-assigned the colors used to the non-erasure and the erasure condition. An overall significance test revealed that the erasure effect obtained across the four studies could very unlikely be explained by "lucky" color assignments to the sub-conditions ($p < .005$). Rather, a true erasure effect based on a non-classical correlation between non-locally separated measurements could more likely explain the likability data. This unusual research strategy testing an effect in each study (Step 1), documenting its robustness across studies (Step 2) and eliminating the confound with permutations (Step 3) constituted an indirect causality test of the non-classical correlation under study.

Maier and Dechamps (2025, p. 21) argued that when attempting a direct causality test within each study it would have created a conflict with the NT axiom. In contrast, their indirect and confound-loaded causal testing strategy according to Maier and Dechamps (2025) ultimately prohibited the occurrence of an "effect and decline" data pattern as proposed by the GQT both within and across studies. Although strong theoretical arguments supported Maier and Dechamps' (2025, p. 21) claim that a direct causality test would lead to an "effect and decline" data pattern in their color studies, empirical evidence obtained from an experimental erasure test without confounding variables was not provided. The present study was performed to fill this gap.

The Present Study

The study presented here was a conceptual replication of the studies described by Maier and Dechamps (2025). The objective color parameter erasure manipulation was almost identical to the original protocol with only a few key differences: Firstly, in the present study the erasure procedure was performed immediately after each participant completed their study instead of after data collection was finished. Secondly, a within-subjects design with 20 color stimulus presentations (10 for each condition) for each participant was used instead of a between-subjects design. Thirdly and most importantly, for each participant all color stimuli were randomly chosen with regard to their hue (0 to 359) and lightness (30% to 70%; saturation was kept constant: 100%). Thus, in contrast to the original design there was no color subset-confound within the experimental manipulation due to the repeated random drawing of colors from all possible options. Due to the absence of the confound a strict causal test of the experimental objective color parameters erasure manipulation on subjective likabil-

ity ratings was established. In line with the predictions of the NT axiom (Römer, 2023b), we predicted an effect and decline pattern when testing a non-classical correlation between measurements of subjective likability of colors and objective assessments of their physical color parameters with such a strict causality testing procedure. This means, initially strong Bayesian evidence for higher likability mean scores in the non-erasure compared to the erasure condition (Bayes Factor$_{10}$ ≥ 10) should be followed by a decline to at least undecided Bayesian evidence or evidence pointing towards the null hypothesis ($BF_{10} \leq 1$; see also Maier & Dechamps, 2025). To test this effect variation across time against random data fluctuations (false-positives), after completion of data collection, a permutation analysis was performed that provided a probability to find such a "strong Bayesian evidence for an effect and later decline" pattern by chance. This strategy to distinguish "effect and decline" from "false-positive" occurrence was originally developed in a similar area of research (e. g., Dechamps et al., 2021; see also Dechamps, 2025).

Methods

Ethical Guidelines

At the beginning of the online experiment, participants were given general information about the study. They were explicitly reminded that participation was voluntary and that their personal data would be handled confidentially. Individuals indicated their willingness to take part by clicking a confirmation button, thereby giving informed consent. All data were stored securely and processed without any identifying information. The study protocol received prior approval from the responsible ethics committee.

Sample

To analyze the data, a Bayesian sequential analysis was conducted. This approach allows data to be collected and analyzed cumulatively until a specific Bayes Factor (*BF*) threshold for either H_1 or H_0 is reached. In this study, since an "effect and decline" data pattern was predicted, a more complex stopping rule was chosen. Initially, in the first part of the sample the appearance of a maximum BF_{10} (BF_{10max}) of at least ≥ 10 for H1 indicating strong evidence for the "effect" was predicted *a priori*. In addition, after reaching a BF_{10max} of this kind a substantial "decline" in Bayesian evidence was predicted, defined as a decrease of the Bayesian evidence to at least undecided evidence (which represents the starting point before data was

collected) or evidence in the direction of H_0 ($BF_{10} \leq 1$). Should such a data pattern occur, data collection will be declared complete and the *BF* curve including final *BF* will be reported. In case no BF_{max} will be reached during data collection, hypothesis testing will be stopped when sufficient evidence for H_0 was established (at least $BF_{01} > 6$ or higher), which would indicate a null finding, i. e. an absence of "effect and decline". Although the study was not preregistered, this testing strategy was verbally agreed on among the researchers prior to data collection and the overall prediction was outlined in Maier & Dechamps (2025, p. 21). The minimal *n* to start Bayesian testing in this study was set to $n = 100$.

Participant recruitment took place through personal contacts and social media, carried out by LMU students as part of experimental courses and bachelor's or master's theses under the supervision of the first author or through the polling company *Prolific*. During data collection, neither the data collectors nor the participants were informed about the specific details of the experimental manipulation or the hypotheses of the study.

In total 6,550 participants took part in the study. Participants were excluded from the data analyses if they indicated that their responses were not reliable ("Hand on heart": *Did you complete this study attentively and conscientiously so that we can use your data*?; $n = 22$) or if they indicated to be color blind ($n = 80$). These exclusion criteria were defined a priori (see also Maier & Dechamps, 2025). The final sample included in the analyses consisted of $N = 6{,}448$ participants.

The sample consisted of 51.30% females, 48.53% males and 0.17% diverse participants. The mean age of participants was 39.23 years ($SD = 13.84$; range: 18–87). Data collection was conducted online via PC or tablet. A participation via smartphone was not allowed. The participants completed the study in German (16.33%), English (75.68%), Spanish (2.83%), French (1.24%) or Portuguese (3.91%) language.

Materials

Color Stimuli

For each participant, twenty colors were randomly selected and randomly assigned to the erasure condition and to the non-erasure condition at the beginning of the experiment. Three standard color parameters from the HSL model (H: hue, S: saturation, L: lightness) were used to define each selected color. The first parameter, hue, covered a range from 0 to 359 degrees on a standard color wheel. A quantum-based random number generator (QRNG) was used to determine the hue of each of the twenty colors by generating random numbers

between 0 and 359. The second parameter, saturation, which typically ranges from 0 to 100%, was kept constant at 100% for all colors used in this study. The third parameter, lightness, describes the brightness of a color and was set to vary between 30% and 70%, ensuring that each hue remained visually distinguishable. The QRNG also specified this value by generating random number between 0 and 40 which were added to 30. Thus, within and across participants the hue values of the colors varied randomly between 0 and 359 degrees, while saturation remained fixed at 100% and lightness varied between 30% and 70%, leading to a set of 14,760 possible colors to be drawn. This procedure introduced controlled uncertainty in both hue and lightness for the experimental color stimuli.

The random selection process was performed automated and server-side which means no inspection or reviewing of the selected stimuli by researchers was necessary. The objective color parameters of the ten colors assigned to the non-erasure condition were stored directly within the result file and were accessible to the experimenters for later inspection. In contrast, the objective color parameters of the ten colors assigned to the erasure condition were not stored and were not documented in any file and remained permanently inaccessible.

During data collection, each participant at each trial was randomly presented with one randomly selected color. Colors were presented as squares with responsive sizing (CSS: min (20rem, 80vw, 80vh)). On typical desktop monitors (1920×1080, browser zoom 100%), this resulted in approximately 320×320 pixels. Exact physical size varied across participants' devices and viewing distances were not controlled. In sum, 20 color trials were presented successively in this way.

Subjective Color Rating

Participants were asked to indicate their subjective liking of the presented color. This was assessed using the following question displayed below the color square: "*How much do you like this color?*" A visual analogue scale (VAS) consisting of a gray horizontal bar was presented beneath the question, ranging from "*not at all*" (0) to "*very much*" (100). The cursor initially appeared at the midpoint of the scale to ensure a neutral starting position.

Procedure

A link to the web-based experiment was sent to participants who could initially choose the language of the study. After giving consent, participants received instructions on the task. A pseudorandom number generator (the randomization module of *jsPsych*) was used to indi-

vidually shuffle the order of all 20 trials covering both within-subjects conditions. At the beginning of each trial, the QRNG provided the randomly selected color parameters defining the color stimulus used in this trial. The condition assignment then determined whether these quantum-generated parameters were permanently stored (non-erasure condition) or discarded after display (erasure condition). This procedure avoided the confound in the Maier and Dechamps' (2025) studies.

It is important to emphasize that in this experiment, neither the participants nor the data collectors had any knowledge of potential color parameter deletions or variations related to data storage. Moreover, participants were not informed about the existence of such conditions.

Twenty color rating trials were presented successively. In each trial, a colored square was displayed at the center of the screen. Above the square, the trial number (X/20) was shown, and below it appeared the visual analogue scale as previously described. Participants were instructed to indicate their subjective liking of the color by moving the slider to the position on the scale that best matched their perception. After participants performed their color rating by pressing the "continue" button, the next trial was presented.

After completing 20 color ratings, participants were asked to provide their age and gender. Next, they were asked whether they were color-blind (yes/no) and whether they had previously participated in a similar study ("*Have you previously participated in a similar study i*n *which colors were evaluated*?"; yes/no). One additional question followed to assess whether the reported data could be included in the analysis ("*Hand on hear*t: *Did you complete this study attentively and conscientiously so that we can use your data*?"; yes/no). Answering the color-blindness item with "yes" and the "hand on heart" item with "no" lead to an exclusion of this participant's data as mentioned above. At the end of each participant's session, the results were stored in an output file that contained only the subjective color data (in the erasure condition) or the subjective and objective color data (in the non-erasure condition).

Design and Statistical Analysis

A within-subjects design was used, consisting of one independent variable with two conditions: The erasure (more precisely in this paradigm: non-storage) of the objective color parameters condition and the non-erasure (storage) of the objective color parameters condition in the final results file.

To test our hypothesis, a one-sided Bayesian *t*-test for dependent samples was conducted.

The analysis examined the mean difference in subjective liking ratings between the non-erasure and erasure conditions. The experimental conditions represented the independent variable, while the mean scores of the subjective liking ratings obtained in each condition served as the dependent variable.

For the Bayesian dependent sample t-test, an uninformative prior was chosen, based on an estimated effect size of $d = 0.1$ with a rather narrow Cauchy distribution centered around zero ($r = 0.1$; i. e., $\delta \sim$ Cauchy[0, 0.1]). In the sequential Bayes factor analysis, participants were included cumulatively in chronological order according to their participation time.

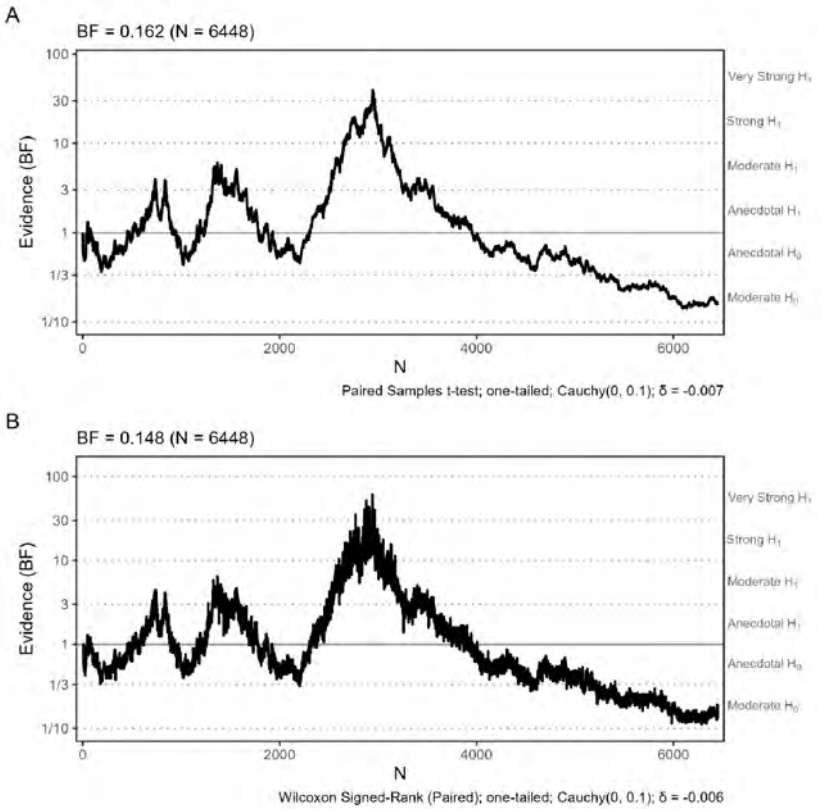
Results

Main Analysis

To test the primary hypothesis that the erasure manipulation of the objective color parameters data will initially lead to strong Bayesian evidence for higher likability means scores in the non-erasure compared to the erasure condition followed by a decline to a $BF_{10} \leq 1$, a one-sided dependent sample Bayesian t-test was conducted with experimental condition (erasure vs. non-erasure) as IV and mean likability rating as DV. The sequential *BF* was monitored during data collection starting at $n = 100$ with additional data being added for each new incoming participant. The Bayesian dependent sample *t*-test (one-tailed) yielded a maximum $BF_{10} = 39.77$ at $n = 2{,}946$, indicating very strong evidence in support of H_1 ($p < .001$; $d_{cohen} = .06$). The mean likability rating score at this n was higher in the non-erasure condition ($M = 57.82$, $SD = 13.95$) than in the erasure condition ($M = 57.19$, $SD = 14.02$). This very strong evidence was then abruptly followed by a decline resulting in evidence pointing in the direction of H_0. Data collection was continued until a final $BF_{01} = 6.16$ was obtained ($p = .29$; non-erasure condition: $M = 57.66$, $SD = 13.46$; erasure condition: $M = 57.59$, $SD = 13.47$; $d_{cohen} = .007$; $N = 6{,}448$). After this $BF_{01} > 6$ was reached, data collection was stopped since it was considered sufficient moderate Bayesian evidence for H_0 (see Figure 1A).

Figure 1

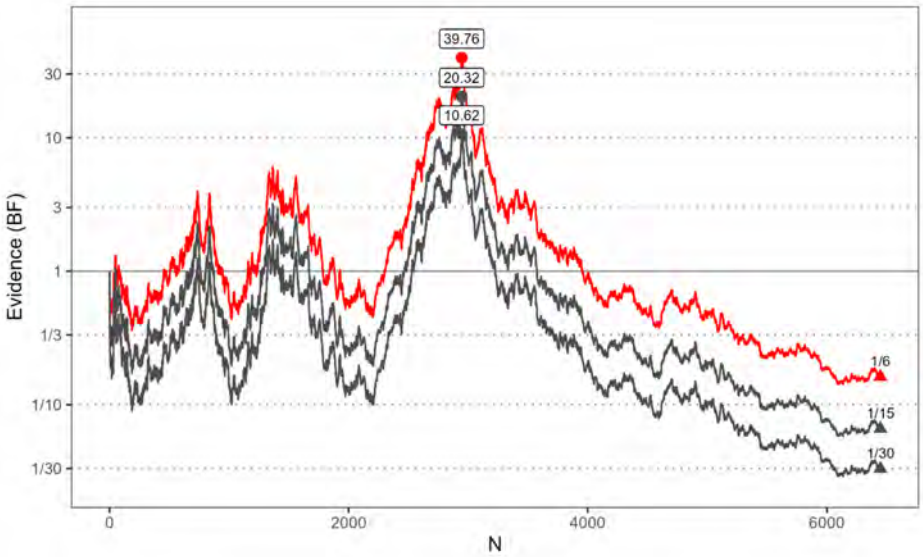
Sequential Bayesian Analyses of the Main Hypothesis Tested With a t-Test (A) and Wilcoxon Signed-Rank Test (B).



In addition, a Bayes factor robustness analysis was performed to explore the impact of the Cauchy prior ($\delta \sim$ Cauchy [0, 0.1]) used in this experiment. The same main analysis was repeated with broader prior widths $r = .2$ and $r = .5$). As can be seen in Figure 2 the "effect and decline" data pattern reaching initially strong Bayesian evidence for H_1 ($BF_{10} > 10$) followed by a remarkable decline could also be found across these wider priors. This indicated that the effect found was robust against the prior chosen.

Figure 2

Robustness Analysis for Different Cauchy Priors: r = .1 (red), r = .2, and r = .5 (both dark grey).



Finally, since the sequential BF of the main analysis showed some unusual fluctuations especially in the first half of the sample which might be caused by extreme outliers in the data, the main analysis was repeated with a non-parametric test (paired Wilcoxon signed-rank test; one-sided). As can be seen in Figure 1B virtually the same sequential BF pattern was found with this test indicating that the original finding cannot be attributed to extreme outlier effects.

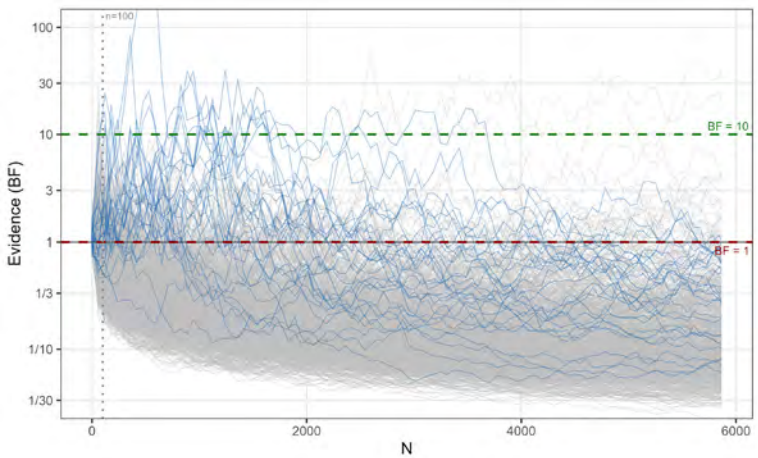
Permutation Analysis

In addition to the primary analysis presented above, an exploratory permutation analysis was conducted to investigate the likelihood to find such an "effect and decline" data pattern by chance. As mentioned above the sequential BF trend found might be considered identical to a false-positive data curve. This study comprises the first attempt of testing macroscopic complementarity between subjective likability of colors and documentation of their objective physical parameters with a strict causality test approach without a confounding variable. Therefore, it was necessary to estimate the probability of initially finding such strong Bayesian evidence for H_1 followed by an extreme decline and finally moderate evidence for H_0. To assess the empirical relevance of the

"effect and decline" pattern found, we first needed to specify what data would align with such a pattern. We decided that "effect and decline" can best be described in Bayesian terms by predicting strong evidence for the effect ($BF_{10} \geq 10$), followed by a complete leveling off of all previously accumulated evidence, therefore returning to the evidential power before data collection or even finding evidence against the existence of an effect ($BF_{10} \leq 1$). The question was how likely this empirical finding was a product of chance occurrence? To calculate this probability a permutation analysis was performed. All rating data obtained in this study from 20 colors times 6,448 participants were randomly reassigned to participants and conditions. One thousand iterations were produced. From these data, 1,000 sequential *BF*s were calculated, which constituted the null distribution. Next, the number of sequential *BF*s was identified that met two criteria: The sequential *BF*s of interest should have reached at least strong Bayesian evidence for H_1 (max $BF_{10} \geq 10$) at any point after the first 100 observations and then should have declined to a final $BF_{10} \leq 1$. Only 2.90% of the null distribution displayed such a sequential *BF*-curve (see Figure 3). Thus, to find a sequential *BF* that exhibits an effect and decline pattern like the one reported in the main analysis by chance was pretty unlikely. Although descriptively the data resemble a false-positive curve of Bayesian evidence later regressing to the true value this interpretation should be rejected. Rather, given the low probability of occurrence by chance the data seem to support a true "effect and decline" interpretation.

Figure 3

Results of the Permutation Analysis Assessing the Likelihood of Obtaining the Effect & Decline Pattern. Sequential Bayes Factors Across 1,000 Permutations. Blue Lines (2.90%) Show the Empirical Pattern of First Exceeding BF = 10 and Subsequently Falling Below BF = 1 After n = 100; Gray Lines (97.10%) Do Not Meet This Criterion.



Note that our empirical finding showed an even stronger decline than just a cancelling out of the previously collected evidence ($BF_{01} > 6$). This is to be expected, since the *BF* will continuously descend in the absence of an effect. The level of additional evidence for a decline can therefore be chosen freely, as long as it is in the direction of H_0.

Discussion

The primary objective of our study was to conceptually replicate the objective color parameter erasure findings on subjective likability ratings originally obtained with a confounded experimental manipulation by Maier and Dechamps (2025) with a strict, unconfounded erasure-dependent causality test. According to the GQT (Atmanspacher et al., 2002; Walach & Römer, 2000, 2011; see also Fach, 2011; Filk & Römer, 2011; Hinterberger & von Stillfried, 2013; Römer, 2023a, 2023b; Lucadou et al., 2007; Walach & von Stillfried, 2011), the proposed non-classical correlations between subjective assessments and objective documentations of color features, with both measurements being non-locally separated from each other (since subjective measurements were performed on the participant level and objective documentation on the experimenter level), are essentially best described as acausally entangled complementary states of observations that must defy causality tests (Lucadou et al., 2007; Römer, 2023b). This implies that strict causality tests will ultimately destroy any acausal complementary relations such as the non-classical correlation studied in this color research, leading to an "effect and decline" data pattern both within or across studies. Maier and Dechamps (2025, p. 21) argued that a purely random assignment of color stimuli to conditions and thus the absence of a "biased colors" confound should lead to an "effect and decline" in an otherwise highly similar study of psychophysical complementarity. In sum, they proposed that according to the NT axiom (Lucadou et al., 2007; Römer, 2023b) such an effect would not be robust.

In their publication they did not provide empirical evidence for this claim which was solely based on the conjectures derived from the GQT. The present study tries to provide the missing evidence by directly testing the specific prediction derived from the NT axiom (Lucadou et al., 2007; Römer, 2023b) with regard to the color-erasure effect. By ensuring fully random selections of color stimuli for each participant, this study established a strict, unconfounded test of the effect of objective data availability on subjective likability rating of color stimuli. The prediction was that strong Bayesian evidence for the original effect, revealing a higher likability mean score in the non-erasure compared to the erasure condition, will be initially found during accumulative data collection. In addition, once this effect would have been documented Bayesian evidence should decline and the evidence should be at least leveled. The results obtained

matched this prediction. The erasure manipulation of objective color parameters on the level of experimenters affected the subjective likability ratings of the participants in the predicted way and initially very strong Bayesian evidence (BF_{10} = 39.77 at n = 2,946) for H_1 was found. This subset of data conceptually replicated the finding reported by Maier and Dechamps (2025). Further, as predicted by the NT axiom, the Bayesian evidence dropped dramatically until a final BF_{01} = 6.16 was found and the study was completed. In addition, the exploratory permutation analysis, which randomly re-assigned colors to participants and conditions revealed that such or a similar remarkable change of Bayesian evidence was unlikely to occur by chance ($p < .05$). This made the interpretation of such a sequential *BF*-curve found with this data as false-positive finding rather unlikely.

Overall, the data obtained from this color rating study in our view supports the conjectures of the GQT (Atmanspacher et al. 2002; Walach & Römer, 2000, 2011; see also Fach, 2011; Filk & Römer, 2011; Hinterberger & von Stillfried, 2013; Römer, 2023a, 2023b; Lucadou et al., 2007; Walach & von Stillfried, 2011) according to which macroscopic complementarity relations might exist between objective and subjective measurements of color features. The "effect and decline" pattern found here highlights two central features of the GQT, namely the non-commutability conjecture and the NT axiom (Römer, 2023b). The non-commutability conjecture predicts that the existence or non-existence of objective measurements about colors non-classically correlates with the subjective likability of the colors. The present findings – at least initially – conceptually replicated those reported by Maier and Dechamps (2025) and thus provided additional evidence for this conjecture. The data also supported the NT axiom, as evidenced by the finding of a decline in the erasure effect. The absence of declines with reduced causality testing (Maier & Dechamps, 2025, Studies 1 to 4) and the appearance of a decline with strong causality testing in this study is fully in line with the predictions of the NT axiom. It should be noted here that the original studies and the present study also differed in their design: The former were performed with a between-subjects design and the later used a within-subject design. We do believe, however, that this difference cannot explain the difference in robustness found between the two projects, since the within-subjects design – initially – clearly produced similar data as those consistently found with the between-subjects designs.

Since the non-classical correlation effect and its decline described here involves interactions between independent measurements performed at the participant level and at the experimenter level and since the results were monitored by the experimenters at least sporadically during the data collection due to the Bayesian testing approach, the possibility of experimenter-psi (e-psi) effects (e.g., Kennedy & Taddonio, 1976) needs to be considered as an alternative explanation of the reported effect. Specifically, the assignment to the experimental condition (erasure vs.

non-erasure condition) was performed by a random process which could have unconsciously been affected by the investigators of the effect under study. This would reflect a mind-matter effect also known as micro-psychokinesis (micro-PK; for an overview see Varvoglis & Bancel, 2015). However, such an e-psi effect would entail a rather complex influence on the data, since for each participant the experimenters would have had to bias the assignment of each specific color to one of the two experimental conditions in a way that the individually most preferred colors were more likely assigned to the non-erasure condition. Since colors were randomly chosen for each trial and each participant, the experimenters would have needed to know or correctly guess what colors the respective participant would prefer. Although in principle possible it would have been extremely complex to perform involving a clairvoyance and micro-PK mixture of some sort. We therefore consider such an e-psi effect rather unlikely.

With regard to an e-psi effect during data monitoring, the PMIR model (Stanford, 1974) suggests that psi results can be biased unconsciously by experimenter's implicit beliefs. It seems therefore possible that the effects reported in our studies and their variations reflected the unconscious beliefs of the investigators coming into acting during data monitoring. Procedurally, the data were not monitored on a regular basis but rather sporadically (the first analysis was performed after 2,000 participants' data have already been collected). Since the sequential BF data before the first observation already displayed some remarkable changes, these and possibly the later curve cannot be attributed to an e-psi-related bias during data monitoring alone. Thus, although these or other psi related effects cannot and might even never fully be ruled out, our conclusion with regard to the effect and decline pattern would be that the erasure interpretation was the most plausible candidate responsible for the results. Nevertheless, more effort needs to be spent in future studies to minimize potential e-psi effects by the involvement of uninformed data analysts.

Conclusion

In sum, previous (Maier & Dechamps, 2025) and the present data suggest that meaningfully connected subjective assessments and objective documentations of color stimuli might best be described as macroscopic complementary relations that take on the form of nonlocal entanglement correlations when subjective and objective measurements are spatio-temporally separated. It seems that when subjectivity is involved, reality cannot be described in a classical-local physical worldview anymore. Rather, the integration of subjective aspects of reality into objective reality descriptions seems to require a whole new perspective. If subjective elements become an integral part of a psychophysical system, then reality becomes observation-dependent, complementary and entangled even on a macroscopic level. This is

because observations or measurements (in this case of colors), whether subjective or objective, non-locally change the system's eigenstates in an entangled manner (Römer, 2023b). Such psychophysically entangled realities might be best characterized by a co-creation of objective and subjective elements due to their acausal entanglement.

Such co-creational effects do not pose a challenge to the so far well documented causal closure of the physical realm (Stoljar, 2024), since they cannot fully be objectified by causal testing approaches. Rather, within strict scientific methods of objectification these non-classical correlations descriptively appear as "false-positives" leaving the causal closure of the objective physical world intact on an epistemic level. Thus, the NT axiom on a meta-level ensures, that a physicalist world view (Stoljar, 2024), with its natural laws governing all physical aspects of reality, is compatible with a co-creational reality foundation as proposed by the GQT, since both can be considered complementary versions of reality descriptions. The physicalist model of reality by its objective documentary nature excludes subjectivity by definition. It could thus be considered a special case of reality established by purely objective effect documentation. Due to the equation of objectification and reality, which is characteristic of physicalism, an epistemically defined objective reality must be devoid of any elements related to subjectivity, including non-physical qualia (Chalmers, 1995, 1996; Lewis, 1929), free will (Kane, 2002; Shariff et al., 2008), and the meaning of life. Any phenomena that do not meet the criterion of strict objectification are thus naturally excluded from "objective reality" by the epistemic strategy that defines "real" phenomena. However, in case subjectivity and its psychophysical relation to the objective realm is addressed, reality descriptions need to be extended as proposed by the GQT, which suggests that these phenomena cannot be strictly causally tested and thus not be fully objectified according to scientific standards. As shown in the work of Maier and Dechamps (2025) and reported here such phenomena can nevertheless be detected and classified as "real" albeit a full objectification procedure cannot be provided in these cases. This research agenda primarily deals with the reduced-objective (but not purely subjective) nature of these phenomena, for which Maier et al. (2022) introduced the term "sobjective." This agenda operates by providing testing strategies which are not fully causal and objective as in standard scientific protocols, but which still might allow to document effects in a convincing, robust and reliable manner. In our view, these less strict causality tests allowed to scientifically document psychophysical complementary relations between subjective and objective elements of reality beyond reasonable doubt and to distinguish them from "false positive" effects (present study) or purely "confounded effects" (Maier & Dechamps, 2025).

In our view, the peaceful co-existence of psychophysical, complementary realities as described by GQT and objective-local realities as proposed by physicalism is established

by their respective scientific documentation strategies. From a physicalist point of view, which only accepts facts when tested with a strict causal objectification approach, the macroscopic complementary relations reported here and by Maier and Dechamps (2025) appear as "false positive" or "confounded" phenomena and would be not considered "real." From a GQT perspective, applying a less stringent yet still compelling testing strategy reveals these complementary relations as equally "real." Whether the physicalist or the GQT reality description is correct depends on which testing and measurement strategies are accepted and defined as scientific. Both worldviews would then be complementary facets of a potentially overarching, undivided reality similar to Bohr's (1928, 1948, 1949) original definition of "complementarity."

Acknowledgments

We would like to thank the students of the Empra Course for their support during data collections.

Finally, we would like to thank Anna-Lena Beinhofer, Marina Drumm and Klara Obretenova for their support in data collection.

Authors' Contributions

MAM proposed the theory and originally designed the study. He drafted the first version of the manuscript. AV helped in data preparation and was actively involved in revising the manuscript. JS helped in data preparation and collection. MCD, programmed the studies, conducted data analyses, helped in data preparation and was actively involved in revising the manuscript.

Declaration of Interests

The Authors declare that there is no conflict of interest.

References

Aspect, A., Dalibard, J., & Roger, G. (1982). Experimental test of Bell's inequalities using time-varying analyzers. *Physical Review Letters, 49*(25), 1804–1807.

Atmanspacher, H., Römer, H., & Walach, H. (2002). Weak quantum theory: Complementarity and entanglement in physics and beyond. *Foundations of Physics, 32*, 379–406. https://doi.org/10.1023/A:1014809312397

Bohr, N. (1928). The quantum postulate and the recent development of atomic theory. *Nature, 121*, 580–590. https://doi.org/10.1038/121580a0

Bohr, N. (1948). On the notion of causality and complementarity. *Dialectica, 2*, 312–319. https://doi.org/10.1111/j.1746-8361.1948.tb00703.x

Bohr, N. (1949). *Discussions with Einstein on epistemological problems in atomic physic*s. Cambridge University Press.

Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2*, 200–219.

Chalmers, D. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.

Dechamps, M.C. (2025). *Meta-analysis of a complete micro-pk database from the LMU Lab: Years 2016–2025*. https://doi.org/10.5281/zenodo.16901945. Retrieved from https://mrzdcmps.github.io/labreport

Dechamps, M.C., Maier, M.A., Pflitsch, M., & Duggan, M. (2021). Observer dependent biases of quantum randomness: Effect stability and replicability. *Journal of Anomalous Experience and Cognition, 1*(1–2), 114–155. https://doi.org/https://doi.org/10.31156/jaex.23205

Descartes, R. (1641). *Meditationes de prima philosophia: Parisiis apud Michaelem soly sub signo of the Phoenicis*.

Fach, W. (2011). Phenomenological aspects of complementarity and entanglement in exceptional human experiences (ExE). *Axiomathes: An International Journal in Ontology & Cognitive Systems, 2*1, 233–247. https://doi.org/10.1007/s10516-010-9143-7

Filk, T., & Römer, H. (2011). Generalised Quantum Theory: Overview and latest developments. *Axiomathes, 21*(2), 211–220. https://doi.org/10.1007/s10516-010-9136-6

Heisenberg, W. (1927). Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik, 43*, 172–198. https://doi.org/10.1007/BF01397280

Hinterberger, T., & von Stillfried, N. (2013). The concept of complementarity and its role in quantum entanglement and generalized entanglement. *Global Philosophy, 23,* 443–459. https://doi.org/10.1007/s10516-012-9187-y

Kane, R. (2002). Free Will, Determinism, and Indeterminism. In H. Atmanspacher & R. Bishop (Eds.), *Between chance and choice: Interdisciplinary perspectives on determinis*m. Imprint Academic.

Kennedy, J.E., & Taddonio, J.L. (1976). Experimenter effects in parapsychological research. *Journal of Parapsychology, 40*(1), 1–33.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly, 64*, 354–61. https://doi.org/10.1111/j.1468-0114.1983.tb00207.x

Lewis, C.I. (1929). *Mind and the world-order: Outline of a theory of knowledg*e. Charles Scribner's Sons.

Lucadou, W.v., Römer, H., & Walach, H. (2007). Synchronistics phenomena as entanglement correlations in Generalized Quantum Theory. *Journal of Consciousness Studies, 14*(4), 50–74.

Maier, M.A., & Dechamps, M.C. (2025). Macroscopic complementary relation between subjective observations and objective measurements of color. *Journal of Anomalistics, 25*(1), 15–60. https://doi.org/10.23793/zfa.2025.015

Maier, M.A., Dechamps, M.C., & Rabeyron, T. (2022). Quantum measurement as pragmatic information transfer: Observer effects on (s)objective reality formation. *Journal of Anomalous Experience and Cognition, 2*(1), 16–48. https://doi.org/10.31156/jaex.23535

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review, 83*, 435–450. doi:10.2307/2183914

Römer, H. (2023a). Twenty years of Generalized Quantum Theory. *Journal of Anomalistics, 23*, 145–154.

Römer, H. (2023b). *Quanten, Komplementarität und Verschränkung in der Lebenswelt: Verallgemeinerte Quantentheorie*. Perspektiven der Anomalistik, Band 7. Lit.

Ruffing, E. (2021). *Einführung in die Geschichte der Philosophie* [Introduction into the History of Philosophy]. 3. Auflage. Brill Schöningh. https://doi.org/10.36198/9783838557595

Shariff, A.F., Schooler, J., & Vohs, K.D. (2008). The hazards of claiming to have solved the hard problem of free will. In J. Baer, J.C. Kaufman, & R.F. Baumeister (Eds), *Are we free? Psychology and free will* (pp. 181–204). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195189636.003.0009

Stanford, R.G. (1974). An experimentally testable model for spontaneous psi events: I. Extrasensory events. *Journal of the American Society for Psychical Research, 68*, 34–57.

Stoljar, D. (2024). Physicalism. In E.N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Spring 2024 Edition), https://plato.stanford.edu/archives/spr2024/entries/physicalism/

Varvoglis, M., & Bancel, P.A. (2015). Micro-psychokinesis. In E. Cardeña, J. Palmer & D. Marcusson-Clavertz (Eds.), *Parapsychology: A handbook for the 21st century* (pp. 266–281). McFarland & Co.

Walach, H., & Römer, H. (2000). Complementarity is a useful concept for consciousness studies: A reminder. *Neuro Endocrinological Letters, 21*(3), 221–232.

Walach, H., & Römer, H. (2011). Generalized entanglement – A nonreductive option for a phenomenologically dualist and ontologically monist view of consciousness. In H. Walach, S. Schmidt, & W.B.

Jonas (Eds.), *Neuroscience, Consciousness and Spirituality* (Vol. 1, pp. 81–95). Springer. https://doi.org/10.1007/978-94-007-2079-4_6

Walach, H., & von Stillfried, N. (2011). Generalized Quantum Theory – Basic idea and general intuition: A background story and overview. *Axiomathes, 21*(2), 185–209. https://doi.org/10.1007/s10516-010-9145-5.

Erweitertes deutsches Abstract

Nicht-klassische Korrelation zwischen subjektiven und objektiven Farbwahrnehmungen: Veränderung des Effekts in Abhängigkeit von seiner empirischen Belegung

In jüngster Zeit berichten Maier und Dechamps (2025) zur Überprüfung der Nicht-Kommutabilitäts-Vermutung der Verallgemeinerten Quantentheorie über makroskopische Komplementaritätsbeziehungen, die empirisch als nicht-klassische Korrelationen zwischen dem subjektiv eingeschätzten Gefallen von Farben und Variationen objektiver Dokumentationen dieser Farben nachgewiesen wurden. Konkret fanden sie höhere mittlere Sympathie-Werte in einer Bedingung, in der objektive Farbparameter nicht gelöscht und in einer Ergebnisdatei gespeichert wurden, im Vergleich zu einer Bedingung, in der diese Parameter gelöscht und für die Experimentatoren unzugänglich waren. Dieser Effekt erwies sich als robust über vier Studien hinweg. Das Design jeder Studie in dieser Versuchsreihe war absichtlich mit einer „Farbbias"-Variable konfundiert. Nur begrenzte Farbsets wurden nach zufälliger Erzeugung gezielt einer der beiden Bedingungen zugeordnet, sodass der Effekt der Löschmanipulation alternativ durch die spezifischen Merkmale der den Bedingungen zugeordneten Farben erklärt werden konnte (die Konfundierung wurde später durch Permutationen statistisch ausgeschlossen). Die Einführung einer Konfundierung erfolgte, um die Auswirkung des NT-Axioms zu vermeiden, das einen strikt kausalen Testansatz verbietet. Gemäß dem NT-Axiom sollte bei strikt kausaler Testung makroskopischer Komplementaritätsbeziehungen vielmehr auf initiale Effektdokumentationen ein Rückgang innerhalb und über Studien hinweg folgen. In der hier vorgestellten Studie wurden die Studien von Maier und Dechamps (2025) durch eine strikte, d. h. unkonfundierte, kausale Teststrategie konzeptuell repliziert, um explizit die vorgeschlagene Auswirkung des NT-Axioms zu testen. In zwei experimentellen Bedingungen (Messwiederholung) wurden objektive Farbparameter entweder gespeichert oder nicht gespeichert. Die Zuordnung von Farben zu Bedingungen erfolgte diesmal auf Probandenebene randomisiert, wodurch ein unkonfundiertes Design sichergestellt wurde. Die aus dem NT-Axiom abgeleitete Vorhersage war, dass unter solchen Umständen der löschungsabhängige Sympathie-Effekt initial gefunden werden sollte und später drastisch zurückgehen würde. Die

Ergebnisse stützten diese Vorhersage: Ein höherer mittlerer Gefallens-Wert wurde initial in der Nicht-Löschungs- im Vergleich zur Löschungsbedingung gefunden mit starker Bayesianischer Evidenz (BF_{10} = 39,77), was die vorherigen Befunde replizierte, nun aber gefolgt von einem Rückgang innerhalb der Studie. Diese Befunde stehen im Einklang mit den Vermutungen der VQT. Konkret deuten sie darauf hin, dass subjektive Bewertungen und objektive Messungen von Farben nicht-kommutabel sind und dass die Daten den Vorhersagen des NT-Axioms folgen, indem sie ein „Effekt-und-Rückgangs"-Datenmuster zeigen, wenn sie strikt kausal getestet werden. Eine zusätzliche Permutationsanalyse zeigte, dass die Wahrscheinlichkeit, ein solches Effekt-und-Rückgangs-Datenmuster zufällig zu finden, sehr gering war ($p < .03$), was darauf hindeutet, dass es kaum als „falsch-positiver" Befund interpretiert werden könnte. Die Auswirkung dieser und der vorherigen Ergebnisse auf die Validität der VQT bei der Beschreibung psychophysischer Beziehungen als nicht-klassische Korrelationen zwischen raum-zeitlich getrennten subjektiven Bewertungen und objektiven Messungen von Farben wird diskutiert.

Schlüsselbegriffe: Verallgemeinerte Quantentheorie, Non-transmission-Axiom, NT-Axiom, makroskopische Komplementarität, makroskopische Non-lokale Verschränkungskorrelation, psychophysische Interaktion, subjektiv-objektive Dualität

Journal of Anomalistics
Zeitschrift für Anomalistik

www.anomalistik.de www.igpp.de

Macroscopic Complementarity Between Subjective and Objective Food Image Assessments

A Conceptual Replication of the Color-Erasure with Confound Effect

Markus A. Maier[a, b, *], Anastasia Vogel[a, b], Moritz C. Dechamps[a]

(a) Department of Psychology, Ludwig-Maximilians-Universität, München, Germany

(b) M. Maier and A. Vogel contributed equally and share first authorship.

[*] Corresponding author: Markus Maier

markus.maier@psy.lmu.de

Abstract – The Generalized Quantum Theory (GQT) proposes that subjective experiences and objective physical measurements constitute macroscopic complementary subsystems related through acausal, non-local entanglement correlations. The non-commutability conjecture of the GQT predicts that measurement within one subsystem modifies the state of the entire system, including the eigenvalues of the complementary subsystem. Maier and Dechamps (2025) provided preliminary evidence for this conjecture by demonstrating that the storage (non-erasure) versus erasure of objective color parameters (hue and lightness) systematically affected subjective likability ratings across four studies (higher subjective likability mean score in the non-erasure compared to the erasure condition). The present research conceptually replicated these color-erasure findings in a different stimulus domain. In four studies (total $N = 647$), participants rated photographs of food dishes on subjective tastiness while the objective food parameters (calorie values and stimulus IDs) were either stored (non-erasure condition) or permanently deleted (erasure condition). As in the color studies, limited stimulus sets were fixedly assigned to conditions within each study, deliberately introducing a stimulus-bias confound to protect the acausal complementary relation from the destructive impact of the non-transmission (NT) axiom. Following a predefined three-step analysis plan, preregistered Study 1 (Step 1) documented a strong erasure effect on

tastiness ratings. Participants evaluated food dishes as more appetizing in the non-erasure compared to the erasure condition (BF_{10} = 1309.22, d = .66), and Studies 2a–2c (Step 2) replicated this effect with three independently drawn stimulus sets (BF_{10}s ranging from 10.40 to 26 × 10^{15}, ds ranging from .26 to .50). In Step 3, preregistered overall analyses including a stratified permutation test (p = .002), a multilevel model (β = 4.72, p = .011), and sensitivity analyses ruled out the biased stimulus assignment alternative. These findings provide cross-domain evidence for the non-commutability conjecture: the documentation of objective stimulus parameters appears to acausally co-relate with subjective evaluations, consistent with macroscopic complementary relations as proposed by the GQT. The implications for substance dualism, physicalism, and dual-aspect monism are discussed.

Keywords: Generalized Quantum Theory, macroscopic complementarity, macroscopic non-local entanglement correlation, non-commutability, erasure paradigm, psychophysical interaction

Introduction

Macroscopic Complementarity and the Psychophysical Problem

The question of how subjective experience relates to the objective physical world – commonly referred to as the psychophysical problem – has occupied philosophers and scientists for centuries (Ruffing, 2021). Within Descartes' (1641) substance dualism, the subjective and objective realms are posited to interact causally, yet numerous scholars have identified a fundamental explanatory gap that such causal accounts fail to bridge (Chalmers, 1995, 1996; Levine, 1983; Nagel, 1974; Shariff et al., 2008). An innovative approach to this long-standing problem has been put forward by the Generalized Quantum Theory (GQT), developed by Walach and Römer (2000, 2011) and Atmanspacher, Römer, and Walach (2002; see also Fach, 2011; Filk & Römer, 2011; Hinterberger & von Stillfried, 2013; Römer, 2023a, 2023b; Lucadou et al., 2007; Walach & von Stillfried, 2011). Rather than postulating causal mechanisms between mind and matter, the GQT proposes that subjective experiences and objective physical measurements constitute macroscopic complementary subsystems whose interrelation is fundamentally acausal in nature.

The concept of complementarity, originally introduced by Bohr (1928) to account for the wave-particle duality in quantum physics, describes a situation in which two observational perspectives are both necessary for a complete description of a phenomenon yet mutually exclusive in their simultaneous determination. The GQT extends this principle beyond the domain governed by Planck's constant, proposing that analogous complementary relationships can

emerge between macroscopic subsystems – specifically between subjective experiential states assessed through a first-person perspective and objective physical states measured through a third-person perspective (Römer, 2023b). When both subsystems are meaningfully connected, for instance through an autonomous and intentional subjective perception of the objective stimulus material, the resulting psychophysical correlation takes on the form of a macroscopic complementary relation. Consequently, the observations derived from each subsystem may be incommensurable: determining the state of one subsystem through measurement changes the state of the entire composite system, including the eigenvalues of the complementary subsystem. This constitutes the non-commutability conjecture of the GQT. If both complementary subsystems are spatiotemporally separated, their measurement outcomes are expected to be related through macroscopic non-local entanglement correlations (Römer, 2023b). Such correlations are acausal in that neither subsystem's measurement outcome can be described as exclusively caused by its complementary counterpart, yet both systematically co-relate.

Maier and Dechamps (2025) recently provided preliminary evidence for the existence of entanglement-based macroscopic complementary relations between subjective assessments and objective measurements of color stimuli, with the data indicating non-commutability between subjective ratings of colors and variations in objective color parameter documentation. The goal of the research presented herein is to re-examine the non-commutability conjecture of the GQT by conceptually replicating these original findings. In this experimental variation, we used different stimulus material involving pictures of food items, from which subjective ratings and variations in the documentation of objective parameters (calorie counts, stimulus IDs) were obtained.

The Problem of Causal Testing and the Non-Transmission Axiom

A central challenge for empirical investigations of macroscopic complementarity lies in the acausal nature of the proposed phenomenon. If such meaning-based psychophysical entanglement correlations could be reliably documented through conventional causal experimental designs, they could in principle be exploited for signal transfer, which would violate the constraints imposed by special relativity (Lucadou et al., 2007; Römer, 2023b). To preclude such paradoxical consequences, the GQT incorporates the non-transmission (NT) axiom, which prohibits the robust causal utilization of macroscopic non-local entanglement correlations (Römer, 2023b). As formalized in the Model of Pragmatic Information (MPI; Lucadou et al., 2007), the NT axiom predicts that when macroscopic complementary relations are subjected to a strict causal experimental test, the resulting data will exhibit an initial effect followed by a sys-

tematic decline in evidential support – a pattern that is phenomenologically indistinguishable from a false-positive finding. This effect-and-decline pattern reflects the progressive destruction of the acausal complementary relation through continued causal objectification.

The implication for empirical research is profound: strict, internally valid causal tests of meaning-based acausal psychophysical phenomena are, according to the GQT, ultimately self-defeating. The very act of maximizing the causal explanatory power of an experimental design eliminates the acausal entanglement that generates the phenomenon under investigation. This theoretical constraint necessitates the development of alternative testing strategies that reduce the directness of the causal test while still permitting convincing empirical documentation of the hypothesized complementary relations (see Maier & Dechamps, 2025).

Empirical Evidence from Color-Erasure Studies

Maier and Dechamps (2025) provided the first systematic empirical evidence for entanglement-based macroscopic complementary relations between meaningfully related subjective and objective aspects of color assessment. In a series of four studies, they investigated whether the storage or deletion (erasure) of objective color parameters, hue and lightness in the HSL model, influenced participants' subjective evaluations of the colors presented. In each study, limited subsets of randomly selected colors were fixedly assigned to either an erasure condition, in which the objective color parameters were omitted from the results file, rendering them irretrievable after completion of data collection, or a non-erasure condition, in which these parameters were stored and remained accessible to the experimenters. Participants rated the colors on visual analogue scales for subjective likability (and also brightness) without any knowledge of the erasure manipulation.

Crucially, the use of limited, fixed color subsets per experimental condition was an intentional design feature rather than a methodological weakness. By restricting each condition to a small set of colors, any observed difference in subjective evaluations could alternatively be attributed to a coincidentally favorable assignment of more appealing colors to one condition (the "biased colors" argument). This deliberate confound reduced the internal validity of each individual study, thereby attenuating the directness of the causal test. We decided to use this strategy to protect the acausal complementary relation from the destructive impact of the NT axiom. If a genuine erasure effect exists, it should appear consistently across studies despite the use of different color sets in each study. Only when the confound is disentangled in a final overall analysis can the true causal nature of the manipulation be retrospectively established.

This three-step research agenda, consisting of (1) documenting an effect in each individual study, (2) demonstrating replicability across studies with different stimulus sets, and (3) ruling out the stimulus-bias alternative through overall analyses, yielded compelling results for the proposed macroscopic complementary relation and the resulting non-commutability between measurements of subjective likability of colors and erasure-manipulated objective color parameter documentation. In all four studies, participants rated colors as more likable when the objective color parameters had been stored (non-erasure) compared to when they had been erased. The Bayesian evidence for this effect was strong to extreme in each study (Study 1: BF_{10} = 22.28, d = .21; Study 2: BF_{10} = 133.2 million, d = .50; Study 3: BF_{10} = 603.16, d = .37; Study 4 [preregistered]: BF_{10} = 10.19, d = .14). The overall permutation analysis across all four studies, which randomly re-assigned colors to erasure conditions, demonstrated that the probability of obtaining the observed pattern of results through coincidentally favorable color assignments alone was very low (p = .005 across all studies; p = .05 for confirmatory Studies 3 and 4 only). A supplementary multilevel model analysis confirmed the robustness of the erasure effect after controlling for color-specific variance (β = 7.46, p = .006, d = .28) and showed that study-level differences, including a lightness-range confound present in Studies 1 and 2, contributed negligible variance. These results provided preliminary support for the non-commutability conjecture of the GQT: the objective documentation of color parameters appears to non-classically correlate with subjective evaluations of the corresponding colors.

Strict Causality Test and the Effect-and-Decline Prediction

While Maier and Dechamps (2025) argued on theoretical grounds that a strict, unconfounded causal test of the erasure manipulation should produce an effect-and-decline pattern in line with the NT axiom, empirical evidence for this specific prediction was not provided in their original study series. This gap was addressed by Maier et al. (2026), who performed a conceptual replication of the original color-erasure paradigm using a maximized-causality testing approach. In their study involving 6,448 participants, 20 color stimuli were individually generated for each participant via a quantum random number generator and randomly assigned to erasure and non-erasure conditions in a within-subjects design. This procedure eliminated the biased-colors confound entirely, as each participant received a unique, fully randomized set of colors. The resulting design constituted a strict causal test of the erasure hypothesis.

The results closely followed the prediction derived from the NT axiom. A sequential Bayesian analysis revealed that the erasure effect on subjective likability initially accumulated very strong evidence in favor of the alternative hypothesis, reaching a maximum Bayes Factor

of BF_{10} = 39.77 at n = 2,946. This initial finding conceptually replicated the results of the confounded studies (higher mean likability ratings in the non-erasure compared to the erasure condition), now under strict causal testing conditions. However, consistent with the NT axiom, the Bayesian evidence subsequently declined dramatically, ultimately yielding moderate evidence for the null hypothesis (BF_{01} = 6.16 at N = 6,448). An exploratory permutation analysis demonstrated that the probability of obtaining such a pronounced effect-and-decline pattern characterized by reaching the BF_{10} > 10 threshold and later declining to at least BF_{10} = 1 by chance was only 2.90%, indicating that a mere false-positive interpretation of the data was unlikely. Robustness analyses across different Cauchy priors and a non-parametric reanalysis confirmed the stability of the observed pattern.

Taken together, the absence of a decline in the confounded studies and its clear presence in the unconfounded study are fully consistent with the predictions of the NT axiom and the GQT more broadly. These findings suggest that subjective evaluations and objective measurements of colors constitute macroscopic complementary subsystems: their entanglement correlation and the resulting non-commutability can be stably documented when the causal directness of the experimental test is reduced but exhibit the characteristic effect-and-decline signature when subjected to a strict causality test.

Reclassifying Color-Erasure Effects as Induced Correlations

In the original presentation of their findings, Maier and Dechamps (2025) classified the color-erasure effects as structural correlations in the terminology of the Pauli-Jung framework (see Atmanspacher & Rickles, 2022). Structural correlations were defined as psychophysical interrelations that lack a normative, collectively shared volitional impulse and thus emerge through individualized, context-specific observations from an originally entangled state. Because color preferences were considered non-normative perceptual processes that do not involve stable, goal-directed intentional constructions, the authors argued that the resulting psychophysical system should be classified as a structural correlation, which should be robust and replicable under appropriate testing conditions (Römer, 2023b).

However, upon further theoretical reflection, we now consider this classification to be overly restrictive. The subjective evaluation of how much one likes a color or, as in the present research, how appetizing a food item appears is not a purely passive registering of external stimuli. Rather, it constitutes an autonomous, experience-based act of perception in which individual preferences, past experiences, and motivational states actively contribute to the evaluative judgment. In other words, intrinsic meaning is present in this psychophysical correlation. In

the language of the Model of Pragmatic Information (Lucadou et al., 2007), such evaluations contain an autonomy-based novelty component ($N > 0$) reflecting the unique, subjective contribution of the perceiver. This novelty component constitutes an individually autonomous contribution to the measurement process that goes beyond passive registration, rendering the underlying psychophysical correlation meaningful.

We therefore reclassify the erasure-dependent effects on subjective likability – both in the color domain and in the food domain investigated here – as induced correlations. In our revised understanding, the defining feature of an induced correlation is not the presence of a collectively shared normative impulse but the presence of an autonomous, experience-based perceptual contribution ($N > 0$) that results in the meaningful formation of a psychophysical reality with pragmatic information greater than zero. This reclassification does not change the fundamental theoretical predictions regarding the erasure effect or the impact of the NT axiom on causal testing. Rather, it highlights the fact that this meaningful psychophysical reality, through its underlying unification of autonomous subjective and deterministic objective aspects, is not entirely objective in nature and thus cannot be fully objectified with a strict scientific causality-testing approach. This revised view also more accurately reflects and aligns with the classification of induced correlations within the broader theoretical Pauli-Jung framework as described by Atmanspacher and Rickles (2022).

The Present Research

The central question motivating the present research is whether the macroscopic complementary relations and the resulting non-commutability between subjective evaluations and objective measurements documented in the domain of color perception (Maier & Dechamps, 2025) can be conceptually replicated with alternative stimulus material. Specifically, we investigated whether the erasure paradigm generalizes to the domain of food perception, where the objective parameters consist of caloric values and stimulus identifiers associated with images of food dishes and the subjective assessment is a rating of how appetizing or tasty the depicted food appears. In this way, the non-commutability conjecture of the GQT is put to a new test.

The food-erasure paradigm shares the essential structural features of the color-erasure paradigm: objective parameters of the presented stimuli (caloric values and stimulus IDs) are either stored in the results file (non-erasure condition) or permanently deleted and rendered inaccessible (erasure condition) after data collection. Participants evaluate the food images on a subjective tastiness scale without any knowledge of the erasure manipulation. As in the color studies, limited subsets of food images are randomly selected from a larger pool and fixedly

assigned to the experimental conditions within each study, thereby introducing the same deliberate confound that reduces the internal validity of each individual study and attenuates the directness of the causal test.

The stimulus material consisted of 100 photographs of attractively arranged food dishes, sourced from the recipe section of a German meal delivery service (HelloFresh). These images were selected to represent a broad variety of cuisines and food types. Each image was associated with an objectively determinable caloric value. In the non-erasure condition, the caloric value together with the stimulus identifier was recorded and stored in a file; in the erasure condition, both pieces of information were not recorded and permanently erased after each study was completed. This manipulation constitutes the independent variable of the present research.

The research program was defined a priori and follows the same data-collection protocol (a first successful Study 1 with three consecutive replications: Studies 2a–2c) and three-step testing agenda established by Maier and Dechamps (2025). In Step 1, the erasure effect on subjective tastiness ratings is documented in a first, preregistered study. In Step 2, the replicability and robustness of the effect are demonstrated across three consecutive independent studies with different, randomly selected stimulus sets. In Step 3, an overall analysis addresses the stimulus-bias confound retrospectively and provides a definitive test of whether the observed effects can be attributed to the erasure manipulation itself rather than to coincidentally favorable stimulus assignments to conditions.

Four studies were conducted, matching the number of studies carried out by Maier and Dechamps (2025) in which the original erasure effect was detected. Study 1 was preregistered and served as the initial test of the food-erasure hypothesis. It employed a within-subjects design with five food images per condition (5 vs. 5), presented across 10 separate sessions with at least two hours between sessions to reduce carryover effects. Each participant rated one image per session, completing 10 sessions in total. Studies 2a, 2b, and 2c (hereafter Studies 2a–2c) served as conceptual replications with a slightly modified design: in each of these studies, six food images per condition (6 vs. 6) were presented within a single session, yielding 12 ratings per participant. Each of the four studies used a different, independently and randomly selected set of food images drawn from the pool of 100 images. Across all studies, the same limited-stimulus-set confound was maintained to protect the acausal nature of the phenomenon, following the rationale established in the color-erasure research.

Hypotheses

Based on the theoretical framework of the GQT proposing non-commutability of measurements derived from meaningfully related macroscopic psychophysical subsystems and based on the empirical precedent established by the color-erasure research (Maier & Dechamps, 2025; Maier et al., 2026), the following hypotheses were formulated.

At the level of individual studies (Steps 1 and 2), it was predicted that participants would rate food images as more appetizing in the non-erasure condition compared to the erasure condition. Specifically, higher mean tastiness scores were expected when the objective caloric values and stimulus identifiers were stored than when they were erased. This prediction was tested in a preregistered manner in Study 1 and in a confirmatory manner, subjected to replicability tests, in Studies 2a–2c. The consistent appearance of this effect across four independent studies with four different sets of randomly selected food images would constitute evidence for Steps 1 (effect documentation) and 2 (replicability) of the research agenda.

At the level of the overall analysis (Step 3), it was predicted that the observed erasure effects could not be attributed to coincidentally favorable stimulus assignments. This hypothesis was operationalized through a preregistered stratified permutation analysis across all four studies. Additionally, a multilevel model analysis was planned in which the erasure manipulation was modeled as a fixed effect and individual stimuli (nested within studies) as random effects, with study included as an additional random factor to account for design differences between Study 1 and Studies 2a–2c. As preregistered, the stratified permutation was expected to yield a significant overall result ($p < .05$), and the fixed effect of the erasure condition in the multilevel model was expected to remain significant after controlling for stimulus-specific and study-level variance. Convergence of both analyses was defined as the strongest evidence against the stimulus-bias explanation. Supplementary sensitivity analyses, including leave-one-stimulus-out analyses and alternative p-value combination methods (Edgington, 1972; Fisher, 1932), were also preregistered to assess the robustness of the overall findings.

Overview of Studies and Three-Step Analysis Plan

Following the three-step analysis plan described above, Step 1 tested the existence of an erasure effect under confound conditions in a first study designed to document an initial erasure effect of objective food image data (erasure vs. non-erasure of calorie values and stimulus IDs) on subjective tastiness rating mean scores. This Study 1 was preregistered on OSF (https://osf.io/zn3sj). Three subsequent replication studies (Studies 2a–2c) were then conducted and are

reported below to test the replicability of the initial effect with new stimulus material. These studies constituted Step 2 of the analysis plan. The replication studies implemented an almost identical experimental protocol and the same main analysis protocol within each study. The only differences between Study 1 and Studies 2a–2c were the number of food-picture stimuli within each set (5 vs. 5 in Study 1 and 6 vs. 6 in Studies 2a–2c), the study-specific random stimulus selection prior to each study's data collection (different sets for different studies), and the number of sessions performed (10 separate sessions in Study 1 and one single session in Studies 2a–2c). This altogether resulted in four studies, which matched the number of studies reported by Maier and Dechamps (2025). The decision to conduct four studies was determined a priori before any data collection commenced. After completion of all four studies, a preregistered overall analysis was performed and is reported below to rule out the potential impact of the "biased stimulus assignment" confound on the data. This preregistration (https://osf.io/ghkr7) was uploaded before any permutation or other combined data scores were known. In sum, the studies performed here constitute a conceptual replication of the color-erasure effects reported by Maier and Dechamps (2025), closely matching the original three-step data-collection and analysis plan.

Study 1

Methods

Ethical Guidelines

Prior to the start of the online experiment, participants were presented with general information about the study. Participation was entirely voluntary, and the data protection regulations were outlined. Informed consent was obtained through an active confirmation via button press. All data were analyzed anonymously. The study was designed, conducted, and analyzed in accordance with the APA guidelines and received approval from the Ethics Committee of the Department of Psychology at LMU Munich.

Sample

The data were analyzed using Bayesian sequential procedures, which enable cumulative evidence updating as new participants are tested and their data are successively incorporated into the dataset. Sampling continued until a prespecified Bayes factor (*BF*) threshold provided strong evidence for H0 or H1. Specifically, an a priori evidential criterion of $BF \geq 10$ (strong

Bayesian evidence) was defined, and the *BF* was tracked throughout data collection. Data collection was terminated as soon as the *BF* reached ≥ 10 in favor of H0 or H1, and the resulting final *BF*s were then reported.

The participants were recruited primarily through private contacts and social media by LMU students in the context of experimental courses under the supervision of the first author. During the data collection period, the experimenters were not privy to the details of the experimental manipulation or the hypotheses of the study.

In total, 111 participants took part in the study. Out of these, $n = 56$ did not complete all 10 sessions and were consequently excluded, as specified in the preregistration. As also specified in the preregistration, participants additionally were to be excluded if they indicated that their responses were not reliable by answering a data integrity item with "No" ("Hand on your Heart: Did you really work on this study attentively and conscientiously so that we can use your data?"; $n = 0$). For 19 participants, this question was not displayed because they switched devices between sessions, which prevented the item from being shown in the follow-up survey. Since the 10-session design with forced pauses between sessions required a high level of commitment, we decided to retain these participants in the dataset, nonetheless. Thus, the final dataset comprised $N = 55$ participants. Data collection was considered complete at this sample size because the Bayesian analyses reached the prespecified evidential criterion ($BF \geq 10$).

The sample consisted of 38 female, 16 male, and 1 participant who identified as diverse. The mean age of the participants was 37.71 years ($SD = 17.17$; range: 18–72). Data were collected online via PC or tablet. Participation via smartphone was not supported, and participants completed the study in German (95%) or English (5%).

Materials

Food Stimuli Used in the Experiment. Prior to data collection, 10 distinct food images were randomly selected from the stimulus pool; five were randomly assigned to the erasure condition and five to the non-erasure condition, serving as the experimental stimuli. All 10 experimental stimuli were presented to the participants in a within-subjects design. After pre-study random drawing and assignment to conditions, the selected images and their assignment to experimental conditions were kept unchanged for the entire study. That is, each participant saw the same sets of five stimuli in each condition. The stimulus pool from which the 10 experimental stimuli were drawn consisted of 100 food images selected by the student researchers from recipe listings on the German HelloFresh website, depicting a broad range of food types. Each image was linked to an objective calorie value. These images and their associated calorie information

served as the objective stimulus attributes that were either stored or erased as a function of experimental condition. Calorie values ranged from 467 to 1,453 kcal (M = 793.29 kcal, SD = 184.19 kcal).

Stimulus selection and condition assignment were automated server-side prior to data collection via a Python script. The script employed a hardware quantum random number generator (QRNG; EasyQuantis, ID Quantique) to directly generate 10 unique integers in the range of 0 to 99 without replacement, each corresponding to a food image in the stimulus pool. The first five indices in the generated sequence were assigned to the non-erasure condition and the remaining five to the erasure condition; since the QRNG produced the sequence in random order, this positional split constituted a random assignment to conditions. During data collection, the presentation order of stimuli with their pre-assigned conditions was randomized individually for each participant. Calorie values and stimulus IDs were stored only for the five stimuli assigned to the non-erasure condition. They were automatically stored directly in the experimental file and were accessible to the researchers for subsequent inspection after data collection. Their image IDs and calorie values were: 36, 706 kcal; 63, 825 kcal; 70, 854 kcal; 86, 947 kcal; 29, 688 kcal. Corresponding information for the five stimuli in the erasure condition was only temporarily recorded and later deleted after study completion. Since the image selection procedure was conducted via an automated Python script, which stored the objective stimulus information temporarily in a separate file, no researcher inspected them at any time. This file containing the objective data was deleted following the end of data collection, ensuring that the calorie values and image IDs could never be identified after the experiment. Consequently, the objective stimulus parameters (calorie values and image IDs) were not subjected to inspection during the selection process, nor while data collection was being performed, nor can they ever be identified afterwards. Performing the erasure manipulation after the completion of the experiment was considered necessary (since the program needed the information to run the study) and was considered sufficient to obtain an erasure effect, since time is not linear in the GQT (see Römer, 2023b).

The use of limited subsets of food images (5 vs. 5) randomly assigned to each experimental condition, with this selection and assignment being kept constant for all participants, established an alternative explanation for any potential effects to be found in this study. Specifically, subsets of food images favoring the confirmation of H1 could have been selected into the respective condition by chance. Although this aspect of the design undermines a clear causal interpretation of the erasure manipulation, it was intentionally established to stabilize the acausal nature of the complementary relation tested in this and all subsequent replication studies. As mentioned above, this alternative explanation will be addressed in preregistered final

overall analyses at the end of the studies section, provided that Steps 1 and 2 were completed successfully.

Subjective Tastiness Ratings. The food stimuli presented to each participant were rated on a 101-point visual analogue scale (VAS) ranging from "not at all" (0) to "very much" (100). Only the endpoint labels were displayed; numeric values were not visible to participants. For each trial, the food image and the rating scale were presented on the same page. Participants were asked, "How delicious does this food look to you?" (German: "Wie lecker sieht dieses Essen für Sie aus?"). Responses were provided using a horizontal grey bar with the labeled endpoints by positioning a cursor on the bar and confirming the response via mouse click. The cursor initially appeared at the midpoint of the scale to ensure a neutral starting position.

Procedure

Participants were provided with a general overview of the study procedure. They were told that the study consisted of 10 sessions in which images had to be evaluated. The 10 experimental stimuli, including their assigned condition, were presented across 10 different experimental sessions, with one image presentation accompanied by the subjective tastiness rating per session. Each subsequent session was accessible only after a minimum two-hour break between sessions. Data collection across separate sessions, rather than presenting 10 trials within a single session, was implemented because this study was part of a larger data-collection project in which one of the concurrent studies required 10 temporally separated sessions. The other studies conducted alongside the study presented here were not related to it and will be reported elsewhere.

Prior to data collection of Study 1, stimulus selection was performed using the QRNG-based procedure described above. During data collection, the presentation order of stimuli across the 10 sessions was randomized individually for each participant using a deterministic seeded pseudorandom number generator (PRNG), producing a fixed permutation of the 10 stimuli for each participant. This ensured that the presentation order was randomized across participants while remaining reproducible for a given participant across sessions.

It is important to note that in this and all subsequent experiments, participants were unaware of the objective data-storage manipulation (erasure vs. non-erasure). In addition, participants could not identify which condition each food image was assigned to, nor were they informed that two different conditions existed.

After a brief introduction to the task, in each session the food picture was displayed in the center of the screen together with the VAS as described above. Participants then rated the perceived tastiness of the food picture ("How delicious does this food look to you?") by moving the

slider to the position on the scale that best reflected their evaluation and confirmed their final rating by pressing the corresponding button.

After completing the ratings across all 10 sessions, participants reported their age and gender. Finally, the data integrity item was presented, asking whether participants had worked attentively and conscientiously so that their data could be used in the analysis. Responses were given by selecting either "Yes" or "No."

Subsequently, the results were saved to a data file that contained only the subjective tastiness data in the erasure condition (with the objective calorie values and image IDs having been permanently deleted), whereas it contained both subjective tastiness data and objective food parameters (image ID and calorie value of the presented food) in the non-erasure condition.

Design and Statistical Analysis

The study used a within-subjects design with one independent variable (IV) with two levels: objective food parameters (picture ID and calorie value of the presented food) were either not available in the final results file (erasure condition) or available (non-erasure condition). The statistical analysis consisted of a one-sided Bayesian paired-samples t-test comparing the erasure and non-erasure conditions. The dependent variable (DV) was the mean score of subjective tastiness ratings for each experimental condition.

For Bayesian inference, we used an uninformed Cauchy prior on the standardized effect size centered at 0 with scale $r = .1$ ($\delta \sim$ Cauchy[0, .1]). Data were accumulated sequentially in chronological order of participation (based on completion time), and Bayes factors were tracked throughout data collection.

All materials, experimental design, hypotheses, and planned statistical analyses reported above were preregistered at OSF (https://osf.io/zn3sj).

Results

To test the directed hypothesis that the mean score of subjective tastiness ratings is higher in the non-erasure than in the erasure condition (non-erasure > erasure), a one-sided Bayesian paired-samples t-test was conducted, with condition (non-erasure vs. erasure) as the within-subject IV and the condition-specific mean score of subjective tastiness ratings as the DV. The Bayesian paired-samples t-test ($N = 55$) yielded a final $BF_{10} = 1309.22$ ($t(54) = 4.88$, $p < .001$), indicating very strong evidence in favor of H1 ($d = .66$). The mean tastiness score was higher in the non-erasure condition ($M = 63.15$, $SD = 14.58$, $n = 55$) than in the erasure condition ($M = 54.99$, $SD = 17.18$, $n = 55$).

In addition to the main analysis, a secondary analysis was preregistered for this study. This involved a permutation analysis with random re-assignments of stimulus-based subjective rating mean scores to experimental conditions to rule out the "biased stimuli" confound. Since this analysis is part of Step 3 of the overall analysis plan, it will be reported in the preregistered overall analysis section below (see sensitivity analyses).

Discussion

The primary objective of Study 1 was to find initial evidence for an erasure-dependent effect of objective food image parameters on subjective tastiness ratings of food images. In the main analysis, we tested whether an experimental manipulation of the documentation of objective food parameters – calorie values and stimulus IDs – affected the subjective evaluation of how appetizing the depicted food appeared. A Bayesian paired-samples t-test with experimental condition (erasure vs. non-erasure of objective food parameters) as IV and mean tastiness rating as DV indicated very strong evidence for H1 (BF_{10} = 1309.22, d = .66). The results demonstrated that participants rated food images as more appetizing when the objective food parameters had been stored (non-erasure condition) compared to when they had been erased.

This finding is consistent with and extends the results reported by Maier and Dechamps (2025), who observed that subjective likability ratings of colors were consistently higher in the non-erasure condition than in the erasure condition across all four of their studies. The present study conceptually replicated this pattern in a different stimulus domain: whereas Maier and Dechamps (2025) manipulated the documentation of objective color parameters (hue and lightness) and measured subjective color likability, the present study manipulated the documentation of objective food parameters (calorie values and stimulus IDs) and measured subjective food tastiness. The convergence of findings across these two distinct stimulus domains suggests that the erasure effect on subjective evaluations may reflect a more general phenomenon rather than being restricted to the specific characteristics of color stimuli.

It is important to note, however, that the results observed in Study 1 can also be attributed to the random selection process employed in choosing the food images used as experimental stimuli. Because only five images were assigned to each condition, it is possible that the images selected for the non-erasure condition happened to depict foods that were inherently rated as more appetizing than those in the erasure condition, resulting in a coincidentally favorable stimulus assignment. This alternative explanation was intentionally built into the design to address the acausal nature of the effect under study. The findings can thus be interpreted in two different ways. They can either be seen as initial evidence supporting the non-commutability

conjecture derived from the GQT, suggesting that the documentation of objective food parameters non-classically correlates with subjective evaluations of the corresponding food images. Alternatively, since there were only few food images per condition, the data can be explained by a coincidentally biased stimulus assignment to conditions. This alternative explanation is currently equally viable and must be addressed by subsequent independent replication attempts and a final overall analysis. Only if the effect found in Study 1 (Step 1) is consistent across different random stimulus selections in independent studies (Step 2) and overall analyses can exclude the biased stimulus argument (Step 3) can the true nature of the effect be determined. Consequently, Studies 2a–2c were conducted to replicate Study 1 with newly selected sets of food images.

Study 2a

The aim of Study 2a was to replicate the effect reported in Study 1, following the logic of the color-study paradigm described by Maier and Dechamps (2025). The replication applied an almost identical experimental protocol with the following differences. First, Study 2a used 12 food images (instead of 10) to match the number of stimuli employed in Studies 3 and 4 of Maier and Dechamps (2025). The 10-stimulus set in Study 1 resulted from the 10-session design, which was central to another study that was also part of the research project, limiting the number of stimuli in Study 1 to 10. This limitation was absent in Study 2a. Second, a newly and randomly selected set of food stimuli, with two new subsets of six images for each condition, was used. Third, instead of 10 separate sessions with one trial each, all 12 trials were presented within a single session. The main objective was to replicate the experimental effect observed in Study 1, using the mean score of subjective tastiness ratings as the dependent variable. Thus, Study 2a (and subsequent Studies 2b and 2c) constituted Step 2 of the three-step analysis plan presented by Maier and Dechamps (2025), in which the confound-loaded erasure effect on subjective tastiness ratings of food images was tested for replicability across three independent studies.

Methods

Ethical Guidelines

Prior to the start of the online experiment, participants were presented with general information about the study. Participation was entirely voluntary, and the data protection regulations were outlined. Informed consent was obtained through an active confirmation via button press.

All data were analyzed anonymously. The study was designed, conducted, and analyzed in accordance with the APA guidelines and received approval from the Ethics Committee of the Department of Psychology at LMU Munich.

Sample

The same Bayesian sequential analysis procedures as those employed in Study 1 were used to analyze the data, with a minimum evidential criterion of $BF \geq 10$ (strong evidence) serving as the stopping rule in the main analysis. If this criterion was met, data collection was terminated and the final *BF*s were subsequently reported.

Participants were recruited primarily through private contacts and social media by LMU students as part of bachelor's thesis projects under the supervision of the first author. During data collection, the experimenters were not privy to the details of the experimental manipulation or the study hypotheses.

In total, 369 participants took part in the study (completers only). As specified a priori, participants were to be excluded if they indicated that their responses were not reliable by answering the "Hand on your Heart" item with "No" ($n = 6$). Thus, the final dataset comprised $N = 363$ participants. Data collection was considered complete at this sample size because the Bayesian analyses reached the prespecified evidential criterion ($BF \geq 10$).

The sample consisted of 211 female, 139 male, and 13 participants who identified as diverse. The mean age of the participants was 30.68 years ($SD = 9.56$; range: 18–100). Data were collected online via PC or tablet. Participation via smartphone was not supported, and participants completed the study in German.

Materials

Food Stimuli Used in the Experiment. In contrast to Study 1, Study 2a used a total of 12 food images derived from the same stimulus pool as those in Study 1. Prior to data collection, 12 pictures were randomly selected and six of those assigned to the erasure condition and six to the non-erasure condition (93, 1,078 kcal; 18, 633 kcal; 9, 591 kcal; 54, 773 kcal; 49, 756 kcal; 14, 616 kcal). Unlike Study 1, in which the QRNG directly generated the stimulus indices, stimulus selection in Study 2a was performed via a QRNG-generated seed. Specifically, a 32-byte random seed was generated by the Quantis USB hardware QRNG and served to the experimental program via a local server endpoint. The client-side experimental script then converted this seed into 12 stimulus indices of the stimulus pool range (0–99). The first six resulting indices were assigned to the non-erasure condition and the remaining six to the erasure condition. The

seed was generated once at server startup prior to data collection and served identically to all participants, ensuring that the same 12 stimuli and condition assignments were used throughout the study. After data collection, the server was shut down, destroying the seed.

As in Study 1, calorie values and stimulus IDs were stored only for the six stimuli assigned to the non-erasure condition. Corresponding information for the six stimuli in the erasure condition was not retained in the output data: on each erasure-condition trial, the actual image path was overwritten with a generic identifier (e. g., "food7") before being committed to the data file. The seed required to reconstruct the original stimulus-to-condition mapping existed only in volatile server memory (RAM) and was never written to persistent storage. Upon server shutdown after data collection, the seed was irrecoverably destroyed. This represents a methodological improvement over the file-based erasure procedure used in Study 1, where the temporary file containing objective stimulus parameters was deleted from disk after data collection. Although that file was never inspected, deleted files can in principle be recovered from persistent storage using forensic tools. The ephemeral-computing approach employed here eliminates this residual recoverability, ensuring that the objective parameters of erasure-condition stimuli are irretrievable by any means after the server process has been terminated.

Subjective Tastiness Ratings. Each food picture presented to participants was rated for subjective tastiness. Ratings were provided on visual analogue scales ranging from 0 to 100, as in Study 1 (see the corresponding section of Study 1 for details).

Procedure, Design, and Statistical Analyses

The procedure, design, and statistical analyses, including the Cauchy prior, were identical to those employed and preregistered in Study 1, with the following exceptions: (a) all 12 trials were presented within a single experimental session (rather than across 10 separate sessions), (b) the presentation order of conditions across trials was randomized for each participant using a shuffle function, which relies on the browser's pseudorandom number generator, and (c) stimulus selection was performed with replacement, meaning that the same image could in principle be drawn more than once across the 12 selections (whereas Study 1 enforced sampling without replacement). The minimum n to initiate Bayesian testing in this and all subsequent studies was set to $n = 100$.

Results

To test the directed hypothesis that the mean score of subjective tastiness ratings is higher in the non-erasure than in the erasure condition (non-erasure > erasure), a one-sided Bayesian

paired-samples *t*-test was conducted, with condition (non-erasure vs. erasure) as the within-subject IV and the condition-specific mean of subjective tastiness ratings as the DV. The Bayesian paired-samples *t*-test (N = 363) yielded a final $BF_{10} = 26 \times 10^{15}$ ($t(362) = -9.61$, $p < .001$), indicating very strong evidence in favor of H1 ($d = .50$). The mean tastiness score was higher in the non-erasure condition ($M = 72.12$, $SD = 16.04$, $n = 363$) than in the erasure condition ($M = 67.03$, $SD = 17.62$, $n = 363$).

Study 2b

The aim of Study 2b was to replicate the effects observed in Studies 1 and 2a. The experimental design followed the procedures of the preceding Study 2a exactly. The only difference was that in Study 2b a newly drawn set of 12 food images was used as stimuli. The random selection of all food stimuli prior to data collection followed the same protocol as in Study 2a, and the identical selection procedure was applied in both experimental conditions. The objective of Study 2b was to replicate the confound-loaded experimental erasure manipulation effect on mean subjective tastiness ratings as the dependent variable.

Methods

Ethical Guidelines

Prior to the start of the online experiment, participants were presented with general information about the study. Participation was entirely voluntary, and the data protection regulations were outlined. Informed consent was obtained through an active confirmation via button press. All data were analyzed anonymously. The study was designed, conducted, and analyzed in accordance with the APA guidelines and received approval from the Ethics Committee of the Department of Psychology at LMU Munich.

Sample

The same Bayesian sequential analysis procedures as those employed in Studies 1 and 2a were used to analyze the data, with a minimum evidential criterion of $BF \geq 10$ (strong evidence) serving as the stopping rule in the main analysis. If this criterion was met, data collection was terminated and the final BFs were subsequently reported.

Participants were recruited primarily through private contacts and social media by LMU students in the context of experimental courses under the supervision of the first author. During

data collection, the experimenters were not privy to the details of the experimental manipulation or the study hypotheses.

In total, 121 participants took part in the study (completers only). As specified a priori, participants were to be excluded if they indicated that their responses were not reliable by answering the "Hand on your Heart" item with "No" ($n = 0$; data not available for $n = 19$). Thus, the final dataset comprised $N = 121$ participants. Data collection was considered complete at this sample size because the Bayesian analyses reached the prespecified evidential criterion ($BF \geq 10$).

The sample consisted of 52 female, 68 male, and 1 participant who identified as diverse. The mean age of the participants was 28.69 years ($SD = 13.75$; range: 18–77). Data were collected online via PC or tablet. Participation via smartphone was not supported, and participants completed the study in German.

Materials

Food Stimuli Used in the Experiment. In line with Study 2a, Study 2b used a total of 12 independently drawn food images. Prior to data collection, six pictures were randomly selected and assigned to the erasure condition and six to the non-erasure condition (9, 591 kcal; 4, 564 kcal; 28, 683 kcal; 76, 870 kcal; 64, 830 kcal; 82, 926 kcal), serving as the experimental stimuli. The random selection procedure followed the same protocol as in Study 2a.

Subjective Tastiness Ratings. Each food picture presented to participants was rated for subjective tastiness. Ratings were provided on visual analogue scales ranging from 0 to 100, as in Study 1 (see the corresponding section of Study 1 for details).

Procedure, Design, and Statistical Analyses

The procedure, design, and statistical analyses, including the Cauchy prior, were identical to those employed in Study 2a.

Results

To test the directed hypothesis that the mean score of subjective tastiness ratings is higher in the non-erasure than in the erasure condition (non-erasure > erasure), a one-sided Bayesian paired-samples t-test was conducted, with condition (non-erasure vs. erasure) as the within-subject IV and the condition-specific mean score of subjective tastiness ratings as the DV. The Bayesian paired-samples t-test ($N = 121$) yielded a final $BF_{10} = 27.23$ ($t(120) = -3.18$, $p = .001$), indicating strong evidence in favor of H1 ($d = .29$). The mean tastiness score was higher in the

non-erasure condition (M = 66.17, SD = 14.24, n = 121) than in the erasure condition (M = 63.38, SD = 15.31, n = 121).

Study 2c

Study 2c was conducted as a final replication test of the confound-loaded experimental erasure manipulation effects on subjective tastiness ratings reported in Studies 1, 2a, and 2b. This study completed the pre-planned series of four studies and constituted the final replicability test of Step 2 within the three-step analysis plan introduced by Maier and Dechamps (2025). The procedure and statistical analyses followed the same protocol as in the previous two studies. Prior to data collection, a newly drawn set of 12 food pictures (six per condition) was randomly selected and used as the experimental stimuli. The dependent variable was the mean tastiness rating score.

Methods

Ethical Guidelines

Prior to the start of the online experiment, participants were presented with general information about the study. Participation was entirely voluntary, and the data protection regulations were outlined. Informed consent was obtained through an active confirmation via button press. All data were analyzed anonymously. The study was designed, conducted, and analyzed in accordance with the APA guidelines and received approval from the Ethics Committee of the Department of Psychology at LMU Munich.

Sample

The same Bayesian sequential analysis procedures as those employed in the preceding studies were used to analyze the data, with a minimum evidential criterion of BF ≥ 10 (strong evidence) serving as the stopping rule in the main analysis. If this criterion was met, data collection was terminated and the final BFs were subsequently reported.

Participants were recruited primarily through private contacts and social media by LMU students as part of bachelor's thesis projects under the supervision of the first author. During data collection, the experimenters were not privy to the details of the experimental manipulation or the study hypotheses.

In total, 111 participants took part in the study (completers only). As specified a priori, participants were to be excluded if they indicated that their responses were not reliable by

answering the "Hand on your Heart" item with "No" (n = 3). Thus, the final dataset comprised N = 108 participants. Data collection was considered complete at this sample size because the Bayesian analyses reached the prespecified evidential criterion ($BF \geq 10$).

The sample consisted of 63 female and 45 male participants. The mean age of the participants was 31.09 years (SD = 16.34; range: 18–80). Data were collected online via PC or tablet. Participation via smartphone was not supported, and participants completed the study in German.

Materials

Food Stimuli Used in the Experiment. In line with Studies 2a and 2b, Study 2c used a total of 12 new randomly selected food images. Prior to data collection, six pictures were randomly selected and assigned to the erasure condition and six to the non-erasure condition (90, 969 kcal; 99, 1,409 kcal; 66, 833 kcal; 80, 895 kcal; 52, 766 kcal; 38, 716 kcal), serving as the experimental stimuli. The random selection procedure followed the same protocol as in the preceding studies.

Subjective Tastiness Ratings. Each food picture presented to participants was rated for subjective tastiness. Ratings were provided on visual analogue scales ranging from 0 to 100, as in Study 1 (see the corresponding section of Study 1 for details).

Procedure, Design, and Statistical Analyses

The procedure, design, and statistical analyses, including the Cauchy prior, were identical to those employed in Studies 2a and 2b.

Results

To test the directed hypothesis that the mean score of subjective tastiness ratings is higher in the non-erasure than in the erasure condition (non-erasure > erasure), a one-sided Bayesian paired-samples t-test was conducted, with condition (non-erasure vs. erasure) as the within-subject IV and the condition-specific mean score of subjective tastiness ratings as the DV. The Bayesian paired-samples t-test (N = 108) yielded a final BF_{10} = 10.40 ($t(107) = -2.75$, p = .004), indicating strong evidence in favor of H1 (d = .26). The mean tastiness score was higher in the non-erasure condition (M = 59.99, SD = 13.79, n = 108) than in the erasure condition (M = 56.24, SD = 15.25, n = 108).

Discussion of Studies 2a–2c

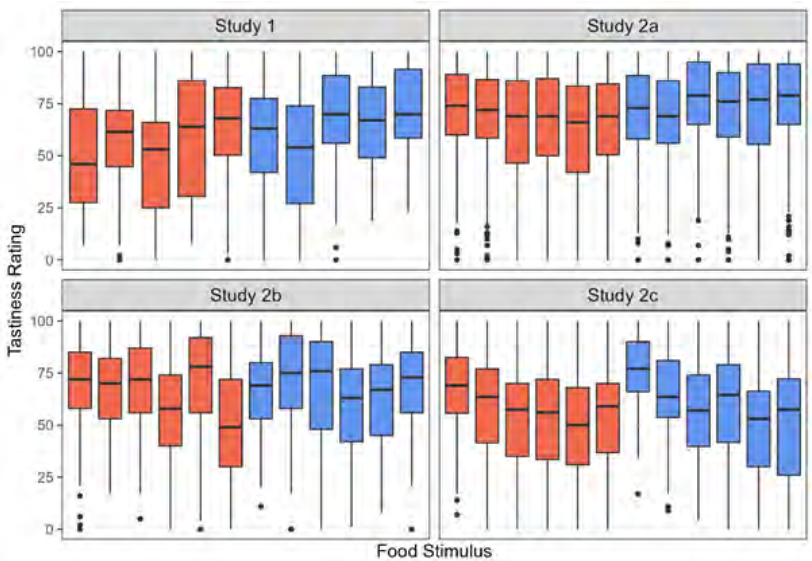
The objective of Studies 2a–2c was to test the replicability of the erasure effect on subjective tastiness ratings initially documented in Study 1. Each of these three studies employed a newly

drawn, independently and randomly selected set of 12 food images (six per condition), while maintaining the same within-subjects design, the same dependent variable (mean tastiness rating), and the same Bayesian sequential analysis protocol. The only design modification relative to Study 1 was the use of an improved erasure protocol, the display of 12 instead of 10 food images, slightly modified sampling and shuffling procedures, and the presentation of all trials within a single session rather than across 10 separate sessions.

In all three studies, the results consistently supported H1: participants rated food images as more appetizing when the objective food parameters (calorie values and stimulus IDs) had been stored (non-erasure condition) compared to when they had been erased (see Fig. 1). The Bayesian evidence was strong to very strong in each study (Study 2a: $BF_{10} = 26 \times 1015$, $d = .50$; Study 2b: $BF_{10} = 27.23$, $d = .29$; Study 2c: $BF_{10} = 10.40$, $d = .26$). Together with the result obtained in Study 1 ($BF_{10} = 1309.22$, $d = .66$), the erasure effect on subjective tastiness ratings was thus documented across four independent studies, each of which used a different set of randomly selected food images. This pattern of findings closely mirrors the results reported by Maier and Dechamps (2025), who observed a consistent erasure effect on subjective likability ratings across four studies with different sets of randomly selected color stimuli.

Figure 1

Mean Ratings of Food Stimuli in the Erasure (Red) and Non-Erasure (Blue) Conditions for Each Study



The convergence between the present food-erasure results and the original color-erasure findings is noteworthy. In both paradigms, subjective evaluations of color likability or food tastiness were systematically higher in the non-erasure condition. The effect was obtained with entirely different stimulus domains (colors vs. food images), different objective parameters (hue and lightness vs. calorie values and stimulus IDs), and different subjective assessment dimensions (likability vs. tastiness). This cross-domain consistency strengthens the interpretation that the observed effects may reflect a general phenomenon of macroscopic complementarity between subjective evaluations and objective measurements, as proposed by the non-commutability conjecture of the GQT.

However, as was the case in each of the individual color-erasure studies reported by Maier and Dechamps (2025), the interpretation of the present findings remains ambiguous at the level of each individual study. Because each study used a limited set of food images fixedly assigned to conditions, the alternative explanation, i.e. that the specific stimulus sets coincidentally favored H1, cannot be ruled out within any single study. Although the probability of such a favorable assignment occurring by chance in four consecutive studies with four independently selected stimulus sets becomes increasingly small, this alternative explanation remains formally viable when each study is considered in isolation.

Our research agenda involved three separate steps. First, the erasure effect needed to be documented in a first study (Step 1). This goal was achieved in Study 1, as reported above. Second, the effect needed to be shown to be replicable across several independent studies with different stimulus material (Step 2). This was also accomplished, as Studies 2a–2c all yielded strong Bayesian evidence for the predicted effect with three different sets of food images. Third, the alternative factor of a biased stimulus assignment argument needs to be ruled out by a final overall analysis (Step 3) to rectify the internal validity violation within the experimental designs, thereby allowing a clear attribution of the effects found so far to the erasure manipulation. This three-step research agenda will next be completed by reporting the results of the preregistered overall analyses.

Overall Analyses for Tastiness: Testing the Impact of Biased Stimulus Assignments

Across the four studies, the experimental manipulation produced a consistent effect on the primary dependent variable, participants' subjective tastiness ratings of food images. This pattern is analogous to the findings reported for the color-based paradigm by Maier and Dechamps (2025) and appeared robust across all four independent samples. However, internal validity is limited because, in each study, the erasure manipulation was intentionally confounded with

stimulus assignment: only a restricted subset of food pictures was randomly allocated to the erasure versus non-erasure condition (Study 1: 5 vs. 5; Studies 2a–2c: 6 vs. 6). As a result, the observed effect cannot be attributed unequivocally to the erasure manipulation; instead, it may reflect a biased assignment of specific food stimuli to experimental conditions. In Step 3 of our research agenda, we therefore subjected this alternative stimulus-assignment explanation to preregistered frequentist statistical tests, specifically a stratified permutation analysis and a multilevel model. To this end, data from all four studies were combined into a single data file (N_{total} = 647).

The stratified permutation analysis and all subsequent analyses reported here were preregistered after data collection was completed. At the time of preregistration, the Bayesian t-test results for the individual studies were already known to the authors; however, neither the stratified permutation analysis nor the multilevel model analysis had been conducted. The preregistration can be accessed at OSF (https://osf.io/ghkr7). Note that in the preregistration the sample size provided for each study in the table was based on the total number of participants including exclusions ($N_{total—with—exclusions}$ = 656). The preregistered analyses reported here, however, were performed on the data without exclusions to be consistent with all data analyses of Steps 1 and 2.

The independent variable was the within-subject erasure manipulation, comprising a non-erasure condition in which objective calorie information and the image ID were stored in the results file, and an erasure condition in which these objective data were permanently deleted and therefore cannot be reconstructed retrospectively. The dependent variable was the subjective tastiness rating, with which participants evaluated each presented food image in terms of how appetizing or tasty the depicted food appeared.

Primary Analysis: Stratified Permutation Test

The purpose of the stratified permutation analysis was to evaluate the hypothesis (H1) that the observed overall erasure effect across the four studies cannot be attributed to a coincidentally favorable stimulus-to-condition assignment (one-sided). In other words, H1 states that the effect reflects a genuine manipulation effect rather than an artifact of biased stimulus allocation. To quantify this, we estimated, via stratified within-study permutations, the probability of obtaining an overall effect at least as large as the observed one if stimulus assignment to erasure condition were effectively random. A small permutation *p*-value (α = .05) would therefore support H1 by indicating that a fortuitous stimulus allocation is unlikely to account for the observed effect.

The procedure permutes stimulus assignments to conditions independently within each study, thereby respecting the nested data structure (stimuli within studies) without assuming that stimuli are exchangeable across studies. For each study, the number of possible assignments of stimuli to conditions is fixed by the study design: Study 1 allows 252 assignments (5 vs. 5), and Studies 2a–2c each allow 924 assignments (6 vs. 6). The full assignment space is the Cartesian product of the study-level assignment sets (approximately 199 million possible combinations). Because evaluating all combinations would be computationally burdensome, we used Monte Carlo sampling and drew 1,000,000 random combinations from this space.

The experimental procedure involved two random components: (a) the selection of 10 (Study 1) or 12 (Studies 2a–2c) stimuli from a larger pool of 100 food images, and (b) the assignment of the selected stimuli to the two experimental conditions. The stratified permutation analysis addresses component (b) by sampling from the full set of possible within-study assignments. Component (a) cannot be tested via permutation because subjective ratings are available only for the stimuli that were actually selected and shown. However, component (a) primarily concerns generalizability rather than confounding. It is addressed by the replication across four independent studies, each using an independently drawn stimulus set. The consistency of the erasure effect across these independently drawn sets, formally evaluated by the stratified permutation analysis, provides evidence that the overall result is not driven by any single fortunate stimulus draw.

First, we computed stimulus-level mean tastiness ratings separately for each study and condition. For each study, we then quantified the study-specific effect as the difference between the mean of the non-erasure stimulus means and the mean of the erasure stimulus means. Because the design was within-subjects and all participants rated all stimuli within a study, stimulus means were based on the same participant sample within each study, so no additional within-study weighting was applied. The condition means (non-erasure vs. erasure) and corresponding differences (non-erasure minus erasure) were as follows: Study 1, $M = 63.2$ vs. 55.1 (difference = 8.14; $n = 55$; 5 vs. 5 stimuli); Study 2a, $M = 72.1$ vs. 67.0 (difference = 5.10; $n = 363$; 6 vs. 6 stimuli); Study 2b, $M = 66.2$ vs. 63.4 (difference = 2.79; $n = 121$; 6 vs. 6 stimuli); and Study 2c, $M = 60.0$ vs. 56.2 (difference = 3.76; $n = 108$; 6 vs. 6 stimuli). These study-specific differences were then aggregated into a single overall test statistic by computing a weighted mean across studies, using each study's participant sample size (n) as the weight, so that larger studies contributed more to the overall estimate. This yielded an observed overall mean difference of 4.70 points (non-erasure minus erasure), which reflects the sample-size-weighted average effect across all four studies.

To assess whether this overall effect could be explained by coincidental stimulus-to-condition assignment, we conducted the stratified permutation test in which stimulus-level mean ratings were randomly reassigned to conditions within each study while keeping the original

number of stimuli per condition fixed. In other words, we permuted the assignment of stimulus means to conditions (rather than participant labels), thereby isolating the potential impact of chance stimulus allocation. For each permutation, we recomputed the study-specific mean differences (non-erasure minus erasure) and then combined them into the overall test statistic using the exact same computation as for the observed effect, namely a sample-size–weighted mean across studies. Based on 1,000,000 Monte Carlo samples from the full assignment space, the probability of obtaining an overall difference at least as large as the observed one in the predicted direction (non-erasure > erasure) was $p = .002$ (Monte Carlo $SE = .000046$), corresponding to 2,148 out of 1,000,000 sampled assignments. Thus, the observed overall effect is unlikely to be attributable to a fortuitous stimulus allocation.

Figure 2

Distribution of the Weighted Mean Difference Across 1,000,000 Stratified (Within-Study) Permutations (Observed Difference Indicated in Red)

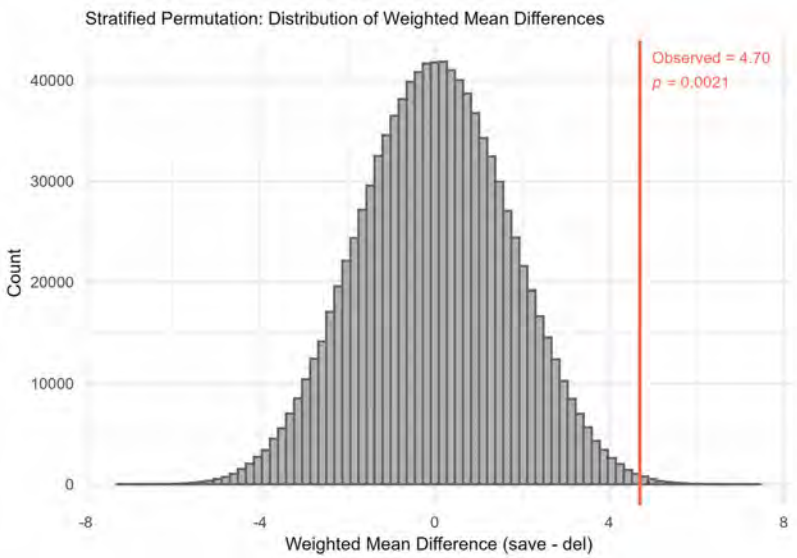


Figure 2 illustrates the permutation null distribution of the weighted overall mean differences and shows that the observed effect falls in the extreme right tail of the distribution.[1]

1 It can be argued that while there is no issue in expanding Bayesian evidence for individual studies

Secondary Analysis: Multilevel Model

Furthermore, we tested whether the erasure effect on subjective tastiness ratings remains significant when controlling for stimulus-specific variance using a multilevel model (MLM; H2, two-sided: mean tastiness ratings differ between non-erasure and erasure). Subjective tastiness ratings served as the DV, and condition (non-erasure vs. erasure) was specified as the IV (fixed effect). To account for stimulus-specific variance, we included random intercepts for stimuli nested within studies. In addition, we included study as a random intercept to model cross-study heterogeneity and to capture design differences between Study 1 and Studies 2a–2c.

The MLM yielded a significant fixed effect of condition, indicating that tastiness ratings were higher in the non-erasure condition than in the erasure condition, $\beta = 4.72$ ($SE = 1.76$), $t(39.11) = 2.68$, $p = .011$, corresponding to a small effect ($d = .19$).

Variance decomposition showed that a small proportion of variance was attributable to stimuli nested within studies (Var = 30.53; $ICC_{stimulus} = .047$) and a similarly small proportion to between-study heterogeneity (Var = 25.54; $ICC_{study} = .039$), whereas most variance remained at the residual level (Var = 590.62). Thus, although some systematic differences between stimuli (and studies) were present, the bulk of variability reflected individual responses rather than stimulus-driven or study-driven effects.

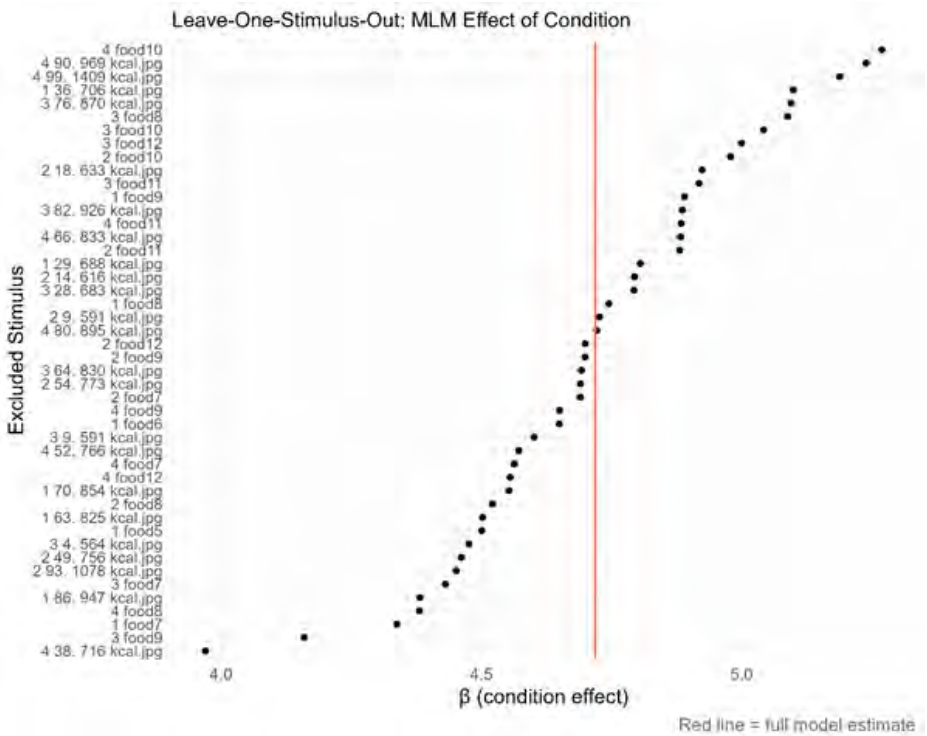
Sensitivity Analyses

Leave-one-stimulus-out Analysis. A leave-one-stimulus-out sensitivity analysis confirmed that the condition effect was stable. Specifically, we refit the MLM while excluding each stimulus in turn and tracked the resulting fixed effect estimate for condition. As shown in Figure 3, the estimated effects clustered closely around the full-model estimate (dashed red line at $\beta = 4.72$) and varied only minimally (β range = 3.97–5.27; all ps < .05). No single stimulus exclusion substantially altered the effect or its significance.

beyond the a priori specified evidence criterion, in case of aggregated data there can be an overrepresentation of favorable data, which might bias a permutation test. We therefore trimmed the experimental data to only include data from each study up until the first time $BF \geq 10$ (for studies 2a to c also with a pre-specified minimum $n > 100$) was met and submitted them to a further permutation analysis. The resulting sample consisted of $n = 355$ participants and showed a weighted mean difference $d = 4.67$ between conditions. Based on 1,000,000 Monte Carlo samples from the full assignment space, the probability of obtaining an overall difference at least as large as the observed one in the predicted direction (non-erasure > erasure) was $p = .008$ (Monte Carlo $SE = .000087$).

Figure 3

Leave-One-Stimulus-Out Sensitivity Analysis of the MLM Condition Effect (Erasure vs. Non-Erasure). Points Show β Estimates; the Red Line Shows the Full-Model Estimate (β = 4.72).



Study-Level Permutation p-Values and Combination Methods. To further assess robustness across studies, we computed exact (exhaustive) within-study permutation p-values for each study separately. The observed differences (non-erasure minus erasure) were positive in all four studies (Study 1: 8.14, $p = .040$; Study 2a: 5.10, $p = .011$; Study 2b: 2.79, $p = .235$; Study 2c: 3.76, $p = .234$), indicating consistent directionality but reduced statistical power in Studies 2b and 2c when considered in isolation. Combining the four study-level *p*-values yielded significant over-

all evidence across two standard combination methods (Edgington: $p = .003$; Fisher: $p = .006$), converging with the preregistered stratified permutation result and supporting the conclusion that the overall effect is unlikely to be explained by fortuitous stimulus-to-condition assignment.

Taken together, these findings suggest that the erasure manipulation effect on tastiness ratings is robust and not driven by specific stimuli.

Discussion

In the overall analyses provided here, Step 3 of the stepwise research agenda was performed. These analyses addressed the problem of a non-ambiguous causal interpretation of the experimental erasure manipulation on tastiness ratings by ruling out the main alternative explanation of "lucky" random stimulus assignments to experimental conditions (stratified permutation analysis and supplementary MLM).

The preregistered stratified permutation analysis provided the most direct test of our central concern. This analysis indicated that the likelihood of finding such an overall experimental erasure effect across studies by chance assignment of stimuli to conditions was very low ($p = .002$). By permuting stimulus assignments to conditions independently within each study and computing a sample-size-weighted overall test statistic across 1,000,000 Monte Carlo samples, this analysis directly addressed the "lucky assignment" alternative explanation while respecting the nested data structure and within-subjects design of the individual studies. In contrast to the pooled permutation approach used by Maier and Dechamps (2025) for the color-erasure data, the stratified approach employed here did not assume exchangeability of stimuli across studies, which is a methodological refinement that more accurately reflects the independent stimulus selection in each study.

The supplementary MLM analysis further supported the robustness of the experimental erasure effect. The results showed that the significant effect of the erasure manipulation ($\beta = 4.72$, $p = .011$) persisted after controlling for stimulus-specific variation within the combined sample. The low ICC values for stimuli (.047) and studies (.039) indicated that most variance came from individual responses rather than systematic stimulus or study effects. The leave-one-stimulus-out sensitivity analysis confirmed that the condition effect remained stable and significant when any single stimulus was excluded, demonstrating that the overall effect was not driven by any particular food image. The sensitivity analyses using alternative p-value combination methods (Edgington: $p = .003$; Fisher: $p = .006$) converged with the stratified permutation result.

In sum, the stratified permutation analysis provided compelling evidence that the observed effects were not due to chance ("lucky") stimulus assignments, while the MLM analysis demonstrated robustness within the observed stimulus set and ruled out study-level confounds. Together, these findings support the conclusion that the erasure manipulation of objective food parameters itself had an effect on subjective tastiness ratings. These analyses rectified the internal validity violation within the experimental settings, making an overall causal interpretation in a retrospective manner possible. They constitute the final Step 3 of our research agenda, which addressed the problem of testing an acausal phenomenon with a scientific-causal approach. Steps 1 and 2 tested the existence of an experimental effect and its replicability across several studies, in each of which an objective data-erasure interpretation on subjective data was confounded with a "biased stimulus assignment to experimental conditions" interpretation. This confound was designed to stabilize the assumed acausal subjective-objective non-commutability relation even under experimental-causal testing conditions within studies. In a third step, the confound was disentangled through overall analyses of the combined dataset. This procedure followed similar recommendations made by Römer (2023b) and Lucadou et al. (2007) and closely parallels the strategy successfully employed by Maier and Dechamps (2025) in the color-erasure paradigm.

As a consequence, these findings lend support to the conjecture derived from the GQT that macroscopic complementary relations exist between objective and subjective measurements of food stimuli. The results indicate that food images are perceived as more appetizing when the objective food parameters are retained compared to a condition in which they are erased. Step 3 analyses thus provided further evidence for the non-commutability conjecture derived from the GQT (Atmanspacher et al., 2002; Walach & Römer, 2000, 2011; see also Fach, 2011; Filk & Römer, 2011; Hinterberger & von Stillfried, 2013; Römer, 2023a, 2023b; Lucadou et al., 2007; Walach & von Stillfried, 2011), which proposed in the present experiments that experimental variations in objective food parameter documentation bias subjective experience of the food stimuli under investigation. Critically, this evidence was obtained in a stimulus domain entirely different from the original color-erasure studies, demonstrating that the non-commutability phenomenon generalizes beyond color perception.

General Discussion

The present research sought to conceptually replicate the macroscopic complementary relations between subjective evaluations and objective measurements documented in the color domain by Maier and Dechamps (2025) using food images as an entirely different class of stimuli. Fol-

lowing the non-commutability conjecture of the GQT, we predicted that erasure of objective food parameters (calorie values and stimulus IDs) would affect subjective tastiness ratings. The three-step research agenda originally developed by Maier and Dechamps (2025) was applied here with the same rationale regarding the deliberate confound and retrospective causality testing and successfully completed across all three steps.

In Steps 1 and 2, the erasure effect on subjective tastiness ratings was detected in Study 1 and successfully replicated across Studies 2a–2c. In all four studies, participants rated food images as more appetizing in the non-erasure condition than in the erasure condition, with strong to very strong Bayesian evidence (Study 1: BF_{10} = 1309.22, d = .66; Study 2a: BF_{10} = 26×10^{15}, d = .50; Study 2b: BF_{10} = 27.23, d = .29; Study 2c: BF_{10} = 10.40, d = .26). In Step 3, the preregistered overall analyses – stratified permutation (p = .002), multilevel model (β = 4.72, p = .011), and sensitivity analyses (Edgington: p = .003; Fisher: p = .006) – ruled out the biased stimulus assignment alternative, leaving the erasure manipulation as the most probable explanation.

The central contribution of the present research lies in the cross-domain generalizability of the erasure effect. Whereas Maier and Dechamps (2025) documented non-commutability between subjective likability and objective color parameters (hue and lightness), the present studies demonstrate the same phenomenon with entirely different stimulus material (food images), different objective parameters (calorie values and stimulus IDs), and a different subjective assessment dimension (tastiness). The consistency of the erasure effect across these two independent stimulus domains substantially strengthens the evidence for the non-commutability conjecture of the GQT. It suggests that the phenomenon is not tied to the specific characteristics of color perception but may reflect a more general property of the psychophysical system – one in which the objective documentation of stimulus parameters systematically co-relates with subjective evaluative judgments.

With respect to the testing strategy, the same considerations raised by Maier and Dechamps (2025) apply. The retrospective overall analyses re-established a causal interpretation of the data without constituting a strict causality test at the study level. We consider this distinction critical given the NT axiom's prohibition of robust causal utilization of macroscopic entanglement correlations (Römer, 2023b; Lucadou et al., 2007). This indirect approach to causality testing, discussed in detail by Maier and Dechamps (2025; see also Walach et al., 2022), appears to circumvent the scientific causality-testing paradox inherent in the GQT. Importantly, the present findings complement the strict causality test conducted by Maier et al. (2026), who demonstrated the predicted effect-and-decline pattern in an unconfounded color-erasure study. Together, confound-protected stability (present research and Maier & Dechamps, 2025) and

NT axiom-driven decline under strict testing (Maier et al., 2026) form a coherent empirical picture fully consistent with the theoretical predictions of the GQT.

The non-local nature of the documented interactions is supported by the fact that participants were unaware of the erasure manipulation and that the objective data were stored or deleted after subjective responses had been recorded, precluding classical-local information transfer. The robustness and replicability of the effect across four studies with independently drawn stimulus sets lends support to the reclassification of these erasure-dependent effects as induced correlations (see the introduction), consistent with the Pauli-Jung mind-matter framework (Atmanspacher & Rickles, 2022). As in the color domain (Maier & Dechamps, 2025), the acausal entanglement correlations underlying the food-erasure effects were not limited by the NT axiom because strict causality tests were avoided at the study level.

Maier and Dechamps (2025) discussed the implications of the color-erasure findings for central positions in the philosophy of mind, arguing that the data challenge both Descartes' substance dualism (Descartes, 1641) and physicalism (Stoljar, 2024), while aligning with dual-aspect monism (Atmanspacher & Rickles, 2022). The present cross-domain replication reinforces these conclusions. The fact that macroscopic complementary relations were now documented in two independent stimulus domains makes it increasingly difficult to dismiss the findings as domain-specific artifacts. If subjective and objective aspects of reality were either ontically separated (substance dualism) or reducible to a single physical substrate (physicalism), the systematic co-relation between erasure of objective parameters and changes in subjective evaluations across colors and food images alike would remain unexplained. Only if subjective experience constitutes a non-physical yet complementary aspect of reality, as proposed by the GQT and dual-aspect monism, can macroscopic complementarity phenomena of the kind documented here be expected to occur (de Broglie, 1958; Römer, 2023b; Zeh, 2012).

Taking these findings together with the convergent evidence from the color-erasure studies (Maier & Dechamps, 2025) and the strict causality test (Maier et al., 2026), a speculative yet coherent picture of the nature of psychophysical reality begins to emerge. After measurement, subjective experience and objective physical reality appear separated and present themselves to the observer on an epistemic level as two distinct, complementary perspectives on the world. However, on an ontological level, both aspects may be fundamentally unified: prior to measurement, they exist as an undifferentiated, acausally entangled whole. The act of scientifically documenting psychophysical relations with a strict causal testing protocol – including direct replications, the gold standard of documenting objectively existing phenomena in the natural sciences – transfers this acausal, entangled reality into a classical, spatiotemporally local, and causally structured psychophysical reality by effectively destroying its true acausal nature. In

this view, causality and spatiotemporal locality are epistemological features that emerge through the act of scientific documentation, whereas the underlying ontological psychophysical reality is acausally connected, complementary, and non-locally entangled. The deliberate use of a confounded experimental design, which implied a reduced objective testing strategy, in the present research may have preserved this acausally connected structure despite the act of phenomenon documentation, allowing the complementary relation to remain intact and the erasure effect to be stably documented. In other words, the present research may have shifted the existing metaphysical boundary, which presently demarcates empirically objectifiable from non-objectifiable phenomena, toward a new boundary that extends the reach of empirical documentation into a domain previously considered non-objectifiable, albeit at the cost of reduced objectifiability.

If we understand experience-based perception and volitional influence as the prototypical manifestations of induced correlations within this framework, then their seemingly unidirectional causal descriptions on the epistemic level – from objective to subjective (perception) or from subjective to objective (volition) – may in fact represent complementary perspectives on an ontologically singular, acausal psychophysical process. What we subjectively prefer, we tend to manifest more readily in the objective, material world and conversely, what is objectively more fully realized tends to be subjectively perceived as more appealing. Experience-based perception and volitional influence would therefore not constitute two separate causal mechanisms but rather two complementary descriptions of the same underlying acausal psychophysical phenomenon, differing only in the direction of the causal narrative imposed upon them. This speculative account, while requiring further theoretical and empirical elaboration, offers a coherent integrative framework for understanding the acausal complementary relations documented across the present and previous studies within the broader context of the GQT and dual-aspect monism.

It should be noted that the results presented here do not provide an exhaustive test of this theoretical framework. While the empirical evidence for the GQT is still accumulating, the cross-domain replication achieved in the present research together with the convergent evidence from the color-erasure studies (Maier & Dechamps, 2025; Maier et al., 2026) suggests a promising outlook for its future empirical evaluation.

Acknowledgments

We would like to thank the students of the two Empra Courses for their support during data collections of Studies 1 and 2b and for developing the design.

Finally, we would like to thank several bachelor students for their support in data collection of Studies 2a and 2c.

Authors' Contributions

MAM proposed the theory and originally co-designed the studies. He drafted the first version of the manuscript. AV co-designed the studies, programmed the studies, helped in data preparation and was actively involved in revising the manuscript. MCD co-designed the studies, programmed the studies, helped in data preparation and was actively involved in revising the manuscript.

Declaration of Interests

The Authors declare that there is no conflict of interest.

References

Atmanspacher, H., & Rickles, D. (2022). *Dual-aspect monism and the deep structure of meaning*. Routledge. https://doi.org/10.4324/9781003270584

Atmanspacher, H., Römer, H., & Walach, H. (2002). Weak quantum theory: Complementarity and entanglement in physics and beyond. *Foundations of Physics, 32*, 379–406. https://doi.org/10.1023/A:1014809312397

Bohr, N. (1928). The quantum postulate and the recent development of atomic theory. *Nature, 121*, 580–590. https://doi.org/10.1038/121580a0

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2*, 200–219.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.

de Broglie, L. (1958). *La théorie de la mesure en mécanique ondulatoir*e. Gauthier-Villars.

Descartes, R. (1641). *Meditationes de prima philosophia*. Parisiis apud Michaelem Soly.

Edgington, E. S. (1972). An additive method for combining probability values from independent experiments. *The Journal of Psychology, 80*(2), 351–363.

Fach, W. (2011). Phenomenological aspects of complementarity and entanglement in exceptional human experiences (ExE). *Axiomathes, 21*, 233–247. https://doi.org/10.1007/s10516-010-9143-7

Filk, T., & Römer, H. (2011). Generalised quantum theory: Overview and latest developments. *Axiomathes, 21*(2), 211–220. https://doi.org/10.1007/s10516-010-9136-6

Fisher, R. A. (1932). *Statistical methods for research workers* (4th ed.). Oliver & Boyd.

Hinterberger, T., & von Stillfried, N. (2013). The concept of complementarity and its role in quantum entanglement and generalized entanglement. *Global Philosophy, 23*, 443–459. https://doi.org/10.1007/s10516-012-9187-y

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly, 6*4, 354–361. https://doi.org/10.1111/j.1468-0114.1983.tb00207.x

Lucadou, W. v., Römer, H., & Walach, H. (2007). Synchronistic phenomena as entanglement correlations in Generalized Quantum Theory. *Journal of Consciousness Studies, 14*(4), 50–74.

Maier, M. A., & Dechamps, M. C. (2025). Macroscopic complementary relation between subjective observations and objective measurements of color. *Journal of Anomalistics, 25*(1), 15–60. https://doi.org/10.23793/zfa.2025.015

Maier, M. A., Dechamps, M. C., & Rabeyron, T. (2022). Quantum measurement as pragmatic information transfer: Observer effects on (s)objective reality formation. *Journal of Anomalous Experience and Cognition, 2*(1), 16–48. https://doi.org/10.31156/jaex.23535

Maier, M. A., Vogel, A., Storch, J., & Dechamps, M. C. (2026). Non-classical correlation between subjective and objective color observations: Change of effect as a function of its empirical documentation. *Journal of Anomalistics*, *26*(1), 14–36. https://doi.org/10.23793/zfa.2026.014

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review, 83*, 435–450. https://doi.org/10.2307/2183914

Römer, H. (2023a). Twenty years of Generalized Quantum Theory. *Journal of Anomalistics, 23*, 145–154.

Römer, H. (2023b). *Quanten, Komplementarität und Verschränkung in der Lebenswelt. Verallgemeinerte Quantentheorie* (Perspektiven der Anomalistik, Bd. 7). Lit Verlag.

Ruffing, E. (2021). *Einführung in die Geschichte der Philosophie* (3rd ed.). Brill Schöningh. https://doi.org/10.36198/9783838557595

Shariff, A. F., Schooler, J., & Vohs, K. D. (2008). The hazards of claiming to have solved the hard problem of free will. In J. Baer, J. C. Kaufman, & R. F. Baumeister (Eds.), *Are we free? Psychology and free will* (pp. 181–204). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195189636.003.0009

Stoljar, D. (2024). Physicalism. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Winter 2024 ed.). Stanford University. https://plato.stanford.edu/archives/win2024/entries/physicalism/

Walach, H., & Römer, H. (2000). Complementarity is a useful concept for consciousness studies: A reminder. *Neuro Endocrinological Letters, 21*(3), 221–232.

Walach, H., & Römer, H. (2011). Generalized entanglement – A nonreductive option for a phenomenologically dualist and ontologically monist view of consciousness. In H. Walach, S. Schmidt, & W. B. Jonas (Eds.), *Neuroscience, consciousness and spirituality* (Vol. 1, pp. 81–95). Springer. https://doi.org/10.1007/978-94-007-2079-4_6

Walach, H., & von Stillfried, N. (2011). Generalized Quantum Theory-Basic idea and general intuition: A background story and overview. *Axiomathes, 21*(2), 185–209. https://doi.org/10.1007/s10516-010-9145-5

Walach, H., Horan, M., Hinterberger, T., & Lucadou, W. v. (2022). Evidence for anomalistic correlations between human behavior and a random event generator: Result of an independent replication of a micro-PK experiment. *Psychology of Consciousness: Theory, Research, and Practice, 9*(2), 173–188. https://doi.org/10.1037/cns0000233

Zeh, H. D. (2012). *The physical basis of the direction of time* (5th ed.). Springer. https://doi.org/10.1007/978-3-540-68001-7

Zusammenfassung

Makroskopische Komplementarität zwischen subjektiven und objektiven Bewertungen von Lebensmittelbildern: Eine konzeptionelle Replikation des „Color-Erasure"-Effekts unter Berücksichtigung von Störeinflüssen

Die Verallgemeinerte Quantentheorie (GQT) postuliert, dass subjektive Erfahrungen und objektive physikalische Messungen makroskopische komplementäre Teilsysteme bilden, die durch akausale, nicht-lokale Verschränkungskorrelationen miteinander verbunden sind. Die Nicht-Kommutativitätsannahme der GQT sagt voraus, dass eine Messung innerhalb eines Teilsystems den Zustand des Gesamtsystems verändert, einschließlich der Eigenwerte des komplementären Teilsystems. Maier und Dechamps (2025) lieferten erste Evidenz für diese Annahme, indem sie zeigten, dass die Speicherung (Nicht-Löschung) versus Löschung objektiver Farbparameter (Farbton und Helligkeit) systematisch die subjektiven Gefallensbewertungen über vier Studien hinweg beeinflusste (höherer subjektiver Gefallens-Mittelwert in der Nicht-Löschungsbedingung im Vergleich zur Löschungsbedingung). Die vorliegende Forschung replizierte diese Farb-Löschungsbefunde konzeptuell in einem anderen Stimulusbereich. In vier Studien (Gesamt-N = 647) bewerteten die Teilnehmenden Fotografien von Lebensmittelgerichten hinsichtlich ihrer subjektiven Schmackhaftigkeit, während die objektiven Lebensmittelparameter (Kalorienwerte und Stimulus-IDs) entweder gespeichert (Nicht-Löschungsbedingung) oder dauerhaft gelöscht wurden (Löschungsbedingung). Wie in den Farbstudien wurden begrenzte

Stimulus-Sets den Bedingungen innerhalb jeder Studie fest zugeordnet, wodurch bewusst ein Stimulus-Bias-Konfund eingeführt wurde, um die akausale Komplementärrelation vor der zerstörerischen Wirkung des Nicht-Transmissions-Axioms (NT-Axiom) zu schützen. Gemäß einem vordefinierten dreistufigen Analyseplan dokumentierte die präregistrierte Studie 1 (Schritt 1) einen starken Löschungseffekt auf die Schmackhaftigkeitsbewertungen. Die Teilnehmenden bewerteten Lebensmittelgerichte in der Nicht-Löschungsbedingung als appetitlicher im Vergleich zur Löschungsbedingung (BF_{10} = 1309,22, d = .66), und die Studien 2a–2c (Schritt 2) replizierten diesen Effekt mit drei unabhängig gezogenen Stimulus-Sets (BF_{10}-Werte von 10,40 bis 26×10^{15}, d-Werte von .26 bis .50). In Schritt 3 schlossen präregistrierte Gesamtanalysen – einschließlich eines stratifizierten Permutationstests (p = .002), eines Mehrebenenmodells (β = 4,72, p = .011) und Sensitivitätsanalysen – die alternative Erklärung einer verzerrten Stimuluszuordnung aus. Diese Befunde liefern domänenübergreifende Evidenz für die Nicht-Kommutativitätsannahme: Die Dokumentation objektiver Stimulusparameter scheint akausal mit subjektiven Bewertungen zu korrelieren, konsistent mit makroskopischen Komplementärrelationen, wie sie von der GQT vorhergesagt werden. Die Implikationen für den Substanzdualismus, den Physikalismus und den Duale Aspekte Monismus werden diskutiert.

Schlüsselbegriffe: Verallgemeinerte Quantentheorie, makroskopische Komplementarität, makroskopische non-lokale Verschränkungskorrelation, Nicht-Kommutativität, Löschungsparadigma, psychophysische Interaktion

Journal of Anomalistics
Zeitschrift für Anomalistik

www.anomalistik.de www.igpp.de

Comments on

(1) M. A. Maier, A. Vogel, J. Storch, & M. C. Dechamps: ***Non-Classical Correlation Between Subjective and Objective Color Observations: Change of Effect as a Function of its Empirical Documentation***

(2) M. A. Maier, A. Vogel, & M. C. Dechamps: ***Macroscopic Complementarity Between Subjective and Objective Food Image Assessments: A Conceptual Replication of the Color-Erasure Effect***

In: ***Journal of Anomalistics*****, *26*(1) (2026), 14–36 and 37–74**

Beyond the Principle of Elusiveness? Analysis of the Experimental Results of the Erasure-Confound Paradigm

Thomas Rabeyron[a, b]

(a) University of Lyon 2 (CRPPC) and Institut Universitaire de France

(b) University of Edinburgh (Koestler Parapsychology Unit)

Thomas.rabeyron@univ-lyon2.fr

Abstract – This commentary analyzes the results of two protocols by Markus Maier's team that focus on the erasure paradigm, which combines two conceptual innovations: the shift of the task from a volitional intention to a perceptual evaluation, and the deliberate introduction of a confound aimed at maintaining the psi effect (what we propose to call the "erasure-confound paradigm"). After situating this work within the context of the limits of the Rhinian paradigm – particularly with regard to the psi paradox and the principle of elusiveness – we discuss the contributions and limitations of this new research paradigm. The analysis highlights the original and promising nature of the results obtained and

situates them within a broader theoretical framework concerning elusiveness aspects of psi phenomena. Several avenues are also opened to determine to what extent the strategy implemented allows for a sustainable circumvention of the principle of elusiveness.

Keywords: Psi, generalized quantum theory, principle of elusiveness, psi paradox, erasure paradigm, decay effect

Background: Reasons for the Failure of the Rhine Paradigm

It is rare for a new paradigm to emerge in psi research. In this regard, the work conducted by Markus Maier and his team – the two articles presented in this issue representing its current evolution – appears to mark a significant break and constitutes one of the most promising research programs in experimental parapsychology. But before addressing the specific content of these two articles, I believe it is necessary to recall the general context in which this work takes on its meaning, as the significance of Markus Maier's proposal can only be fully grasped in light of the difficulties faced by experimental parapsychology for over a century.

Research on psi phenomena began with the collection of accounts reported by individuals who had experienced "extraordinary" events, which contemporary literature now refers to mainly as "anomalous" or "exceptional" experiences (Cardeña et al., 2014; Rabeyron, 2020). The goal was then to demonstrate in the laboratory, under controlled conditions, the existence of "psi interactions" that characterize some of these experiences and that seem to transcend the usual laws of space, time, and causality. This experimental approach, based in particular on statistics, was developed beginning in the 1930s by J. B. Rhine (1934/1973) at Duke University, giving rise to an experimental paradigm aimed at demonstrating the existence of psi.

Since its inception, this paradigm has produced statistically significant effects, though with modest effect sizes – most often ranging between 0.10 and 0.20 – as reported by a number of meta-analyses (Cardeña, 2018; Mossbridge et al., 2012; Storm et al., 2010). These studies have given rise to recurrent controversies regarding the nature of these effects, controversies fueled by the absence of a theoretical framework capable of explaining the "substrate" from which such interactions might arise. Researchers subsequently applied this same experimental methodology to new paradigms – *remote viewing*, *Ganzfeld*, premonition, retrocausality, etc. (see, e. g., Bem, 2011, and the meta-analysis by Bem et al., 2016) – without, however, succeeding in convincing the scientific community of the existence of psi or in truly producing scientific knowledge on a cumulative model. In light of the philosophy of science as conceived by Lakatos (1970), it then becomes difficult to consider this research program as "progressive": it fails both to formulate strong theoretical hypotheses and to accumulate new knowledge according to the usual criteria of testability and falsifiability. Rhinian-oriented experimental parapsychology

remains, so to speak, indefinitely trapped in its initial stage of demonstrating the existence of psi and thus fails, so far, to become a mature scientific discipline.

In an article (Rabeyron, 2020) and two recent books (Rabeyron, 2023, 2026), I have proposed identifying the epistemological causes of this situation through the "psi paradox." The argument can be summarized as follows: psi phenomena arise and develop within an "entanglement" or "communion" between a subject and its environment; yet scientific activity rests on a radical separation of subject and object, a separation that constitutes precisely the condition of possibility for objective measurement. Attempting to demonstrate the existence of psi interactions within the standard scientific framework thus amounts to invalidating the very conditions of its demonstration. The reciprocal contamination between the scientist and their object of study – which is here constitutive of psi – thus prevents any stable accumulation of knowledge, since we are not able to determine whether the effects obtained correspond to factors external to the experimenter or whether, on the contrary, they are produced by the experimenter's beliefs, expectations, and intentions within the experimental setting.

Another factor at the root of these reproducibility difficulties stems from what I have proposed to call the "principle of elusiveness" (Rabeyron, 2023, 2026). Well known to researchers studying spontaneous phenomena such as poltergeists (Roll, 1977; Tierney, 2012), this principle refers to the fact that psi phenomena tend to reveal themselves by concealing themselves. Their emergence thus seems inseparable from their elusive nature. This principle constitutes a fundamental function related to the very origin of psi interactions: they behave like fireflies that can no longer be distinguished once they are directly illuminated (Rabeyron, 2023). Similarly, psi phenomena can emerge and survive only in the "shadows," and any attempt to explicitly objectify them leads to their disappearance. This observation is also central to the *Model of Pragmatic Information* (MPI) and is known as the "Non-Transmission Axiom" (NT axiom), which posits that psi cannot be used to transmit information, explaining its characteristic elusiveness and difficulty in replication (Lucadou et al., 2007).

Toward a New Research Program

These two factors (the psi paradox and the principle of elusiveness) might explain the difficulties and dead ends inherent in the experimental paradigm when applied to the study of psi. But while it is one thing to diagnose the failure of a paradigm, it is quite another to propose a new one that allows for moving beyond the difficulties encountered. To achieve this, it seems necessary to understand that psi phenomena belong to a different ontological realm than those studied by the "normal" sciences – a fact that a number of philosophers had already intuitively perceived. Plato, with his allegory of the cave, had already sketched out an irreducible duality

between the apparent order of the world and its ontological foundation. Spinoza (1677/1996), for his part, proposed distinguishing the "divine substance" from its "attributes" or "modes" as we can apprehend them. It was then Schopenhauer (1819/1969, 1851/1974), in the wake of Kantian criticism, who developed the idea of a "will" underlying the phenomenal world, of which psi would constitute a form of expression. More recently, the Pauli-Jung Conjecture (Pauli, 1954/1994) and physicist David Bohm's (1980) theory of the "implicate order" have clarified the distinction between two orders of reality, the most primordial of which is the locus of the emergence of psi interactions.

Atmanspacher (2003; Atmanspacher & Rickles, 2022) has provided a contemporary model following these ideas – in particular the Pauli-Jung Conjecture – through a "decompositional dual-aspect monism" (see also Atmanspacher, 2020; Atmanspacher & Fach, 2013, for related extensions). In this model, an "ontic order" is decomposed into a first-person subjective experience and a third-person-apprehendable external world, with these two registers (subjective and objective) maintaining relationships of complementarity and incompatibility between them. Such an effort at formalization also gave rise to the *Generalized Quantum Theory* (GQT) (Atmanspacher et al., 2002; Filk & Römer, 2011; Römer, 2023; Walach & Römer, 2000, 2011; Walach & von Stillfried, 2011), which proposes a formal framework inspired by quantum mechanics to model psi interactions.[1] The question that then arises is that of the experimental translation of such a model.

As mentioned earlier, Lucadou et al. (2007) developed the *Model of Pragmatic Information* (MPI), which incorporates these principles of elusiveness through the non-transmission axiom (NT axiom): they assume that acausal correlations cannot be used causally, otherwise they would violate the constraints imposed by special relativity. This model gave rise to an ingenious research program called the "correlation matrix method" (CMM; Lucadou, 2015; Tierney, Watt, & Flores, 2018; Walach et al., 2022). This paradigm aims to allow sufficient "freedom" for psi effects to persist within the system while moving in the matrix, thereby circumventing the NT axiom. It is thus assumed that a significant number of correlations can persist within a correlation matrix, escaping the constraint of direct documentation. Nevertheless, despite promising initial results, this paradigm appears to have been overtaken by the principle of elusiveness and has gradually led to non-significant results (Walach et al., 2022). The research program developed by Markus Maier and his collaborators builds on this research and aims to find an alternative solution to improve the reproducibility of psi interactions while circumventing the principle of elusiveness.

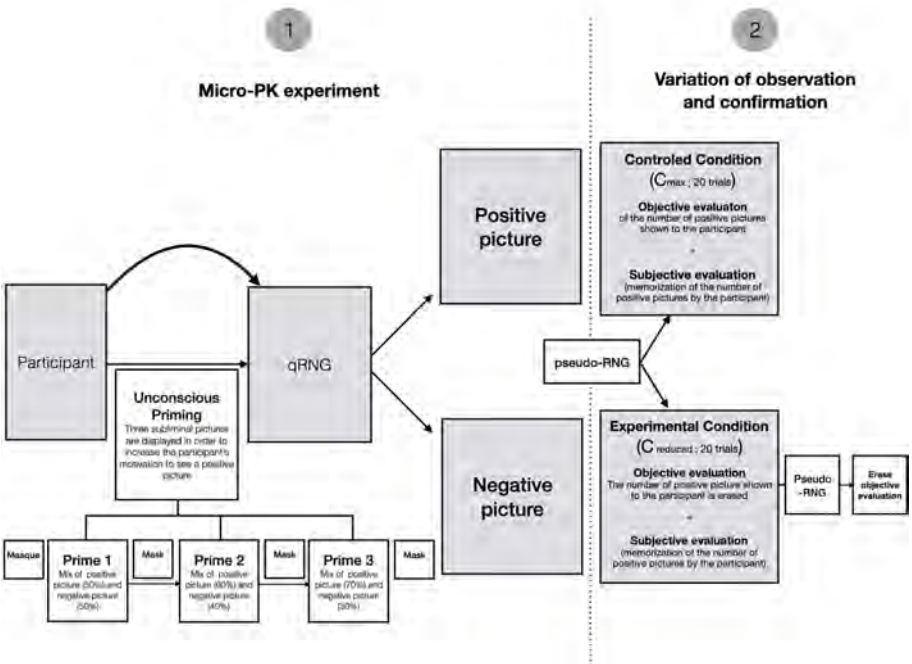
1 In my own work, based in particular on observations in the field of exceptional experiences (Rabeyron, 2020), I likewise arrived at the conclusion that dual-aspect monism seemed to be the most coherent epistemological model for accounting for psi interactions, but only on the condition that the principle of elusiveness be incorporated into it (Rabeyron, 2023).

Protocol Based on a Micro-PK Task and Data Erasure

The program by Markus Maier and his team stems from an intuition: if the observation, documentation, and measurement of psi effects tend to make them disappear, then an asymmetry in documentation between conditions should be introduced into the experimental setting – that is, the degree of objectification of the collected data should be modulated to determine the impact of this parameter. The first implementation of this idea (Maier & Dechamps, 2018; Dechamps et al., 2021; Maier et al., 2022) – in which I participated and which was the result of theoretical exchanges with Markus Maier and his team over the past decade – was based on a micro-PK task in which part of the data was deleted and the other part retained (see Figure 1). The hypothesis was that erasing the data would reduce the degree of objectivity of the observed effects, which we could then verify through a difference in effect size between the two conditions ("erasure" vs. "non-erasure").

Figure 1

Micro-PK and Erasure Experiment (Maier et al., 2022)



I will not go into the details of this protocol and will simply note that, following initial encouraging results, this "erasure paradigm," much like the Correlation Matrix Method, encountered unstable effects that prevented true reproducibility.

Protocol Based on Color Assessment and Data Erasure

It was in response to this "failure" that this new research program was designed, based on a novel conceptual innovation. In a classic parapsychology experiment, participants are usually asked to act on the apparatus according to their intention (for example, to influence an RNG or to guess a target). Maier and Dechamps (2025) introduced a perceptual task based on the aesthetic evaluation of a stimulus. Theoretically, and according to the classification proposed by Atmanspacher and Rickles (2022), the protocol evaluates a "structural correlation" rather than the "induced correlations" typically studied in parapsychology. In the former case, the participant perceives information; in the latter, they attempt to influence the system.[2]

Specifically, participants are asked to subjectively evaluate a color (a colored square displayed on the screen whose hue and brightness vary randomly). Each participant then indicates, on a visual analog scale, how much they like this color. Out of a total of 20 trials, half are stored in a file ("non-erasure" condition), while the other half are permanently deleted ("erasure" condition). In the deletion condition, only the participant's subjective evaluation is retained, without the ability to determine the corresponding color. The central hypothesis is therefore that the storage of the data will have an impact on the subjective evaluation of colors and that it will thus be possible to demonstrate a significant difference between participants' responses depending on whether the objective data related to the images were stored or not.

An initial article by Maier and Dechamps (2025) reported on their theoretical explorations and the variability of the early results obtained using this approach (for example, they initially studied the evaluation of brightness rather than the aesthetic dimension of colors). Maier thus gradually arrived at the idea of introducing a new parameter that may seem counterintuitive but is in fact particularly ingenious for circumventing the problem of elusiveness. This parameter consists of including a confound in the experimental design, which takes the following form

2 Maier, Vogel, and Dechamps (2026), however, revisit this distinction in their second article and propose that we consider the blurring effects as perceptually induced correlations, insofar as subjective evaluation involves an autonomous contribution from the subject and is not merely a passive recording. In reality, such correlations are already used in other experiments, but in a slightly different form. For example, in retro-priming tasks (Bem, 2011), the participant is also asked to perceptually evaluate an image based on their preference (is it positive or negative?).

here: the colors assigned to the two conditions ("erasure" vs. "non-erasure") are the same in an experiment, so that an alternative explanation to a psi effect is conceivable. The effect could result from the fact that the colors selected randomly in one condition might be preferred by participants due to their characteristics. This confound therefore does not allow to determine whether the observed effect is a causal effect or an acausal (psi) effect. Maier and Dechamps (2025) replicated this experiment several times and were able to obtain a significant effect that appeared to be replicable. But how can we then be sure that this effect is a psi effect and not the consequence of the confound described above? Maier and Dechamps (2025) then subsequently used a statistical permutation method that allowed them to rule out this confound and confirm that the effect is not the consequence of the confound.

Empirical Demonstration of the "Effect and Decline" in the Absence of Confound

The first article presented in this issue (Maier et al., 2026) is part of this research program but serves to confirm a hypothesis underlying the research described previously, namely that the absence of confound should lead to a causal prediction and thus eliminate the psi effect. This research therefore aimed to confirm retrospectively, in an empirical manner, the reasons for introducing a confound into the protocol presented previously. In this study, conducted with 6,448 participants, the authors implemented a design in which the 20 colors are, for each participant, individually selected at random via a quantum random number generator (QRNG), thereby eliminating any confound (the colors vary each time, and thus the variation in results between the "erasure" and "non-erasure" conditions cannot stem from the fact that, by chance, the colors in one condition are preferred by the participants).

The prediction derived from the NT axiom is therefore that, under these conditions, the psi effect – which takes the form of a difference in effect size between the "erasure" and "non-erasure" conditions – should initially emerge before declining (as observed in Correlation Matrix Method experiments). The results strikingly confirm this prediction: the Bayes factor reaches a maximum of $BF_{10} = 39.77$ at $n = 2{,}946$ (very strong evidence in favor of H_1), before dropping sharply to a $BF_{01} = 6.16$ (moderate evidence in favor of H_0). An exploratory permutation analysis shows that the probability of obtaining such an "effect-and-decline" pattern is only 2.90%. This result thus appears to provide empirical confirmation of a prediction that had previously remained largely conjectural. Maier and his team thus bridge the gap that was missing from their demonstration: they first produced consistent results across four studies using the erasure-confound paradigm, then replicated the same protocol, this time without confound, to show that it is precisely the *presence* of the confound that allows the psi effect to be maintained in the system.

Protocol Based on Taste Assessment and Data Erasure

The second article revisits the research protocol based on subjective color assessment but this time modifies the stimulus (Maier et al., 2026): participants are now asked to evaluate photographs of food dishes ("*How delicious does this food look to you?*"). The objective data stored in the "non-erasure" condition correspond to the dish's identifier and its associated caloric value. For each study, a subset of photographs is randomly selected using a quantum random number generator: 10 images for Study 1 (5 vs. 5, presented over 10 sessions spaced at least two hours apart) and 12 images in each of the other three studies 2a-2c (6 vs. 6, presented in a single session), with each study based on an independent selection of photographs. Participants' ratings are collected on a visual analog scale ranging from 0 to 100. The article rigorously follows the previous methodology, which has proven effective for color evaluation: (1) initial documentation of an effect; (2) reproduction of the effect while introducing a confound; (3) *a posteriori* elimination of the confound through stratified permutations. From a theoretical standpoint, confirming the effect with a new type of stimulus serves to highlight its potential robustness. It should also be noted that preference for a dish likely involves a greater emotional investment than that related to a color and therefore yields greater variability in the ratings (however, the reasons for this choice of stimuli are not explicitly stated in the article).

Four studies were thus conducted with 647 participants: an initial study (1) yielding a BF_{10} = 1,309.22 (d = 0.66), followed by three other studies (2a; 2b; 2c), whose protocol – slightly modified from Study 1 – remained identical, even though each new study used a new set of stimuli randomly drawn from the same pool of 100 images. The effect was then replicated with varying magnitudes but always in the predicted direction: Study 2a, d = 0.50; Study 2b, d = 0.29; Study 2c, d = 0.26. We must then rule out the confound which, as with the color experiment, stems from the fact that the same photographs are used in both conditions – erasure and non-erasure – for all participants. The permutation test confirmed that the effect cannot be explained by random pairing of stimuli with conditions. This analysis yields p = 0.002, a robust fixed effect in a multilevel model (β = 4.72, p = 0.011), as well as converging sensitivity analyses (Edgington: p = 0.003; Fisher: p = 0.006), ruling out the hypothesis of a confound induced by the random selection of images. These results thus confirm the findings from the experiment conducted with colors and suggest that the observed psi effect is sufficiently stable and robust to persist despite repeated experiments.

Theoretical Comments: Can the Principle of Elusiveness Truly Be Circumvented?

As noted in the introduction, this research program appears particularly innovative and seems to be capable of truly offering a potential solution to the reproducibility difficulties observed for over a century in experimental parapsychology. Nevertheless, I believe it is important to highlight several fundamental theoretical issues in order to put these preliminary results into perspective, as they require further confirmation beyond their innovative nature.

Explanation of the Preference Effect in the Non-masked Condition

The empirical result of these two articles can be summarized as follows: *when the objective parameters of the stimuli are recorded (colors or photographs of food dishes), participants tend to evaluate these stimuli more favorably than when they are obscured*. Demonstrating the effect is essential, but the direction of the observed effect is not explained. But why does recording lead to a more favorable evaluation and not the opposite? Why is the "non-erasure" condition systematically associated with higher evaluation scores? Why does storing the parameters of a photograph of a food dish result in a person preferring that dish?

We thus observe an entanglement, a "co-incidence," between two parameters, without being able to truly determine their causal relationship. This also raises a more fundamental theoretical problem, already identified in the parapsychological literature by the *Decision Augmentation Theory* (DAT; May et al., 1995). According to this theory, micro-PK effects are actually precognition effects: the experimenter anticipates and unconsciously chooses the "right moment" to select the data and begin the experiment, which then results in an effect that appears to be an interaction between the participants and the RNG. The same problem arises in the experiments conducted by Maier and his team, which creates uncertainty regarding the nature of the effect: does it truly originate from the participants, or does it originate from the experimenters? Is it an effect of erasure, or is it an effect of precognition or micro-psychokinesis? In the specific case of this experiment, this translates into the following questions: Is it because the data will be stored that the participants will appreciate certain stimuli more? Or, conversely, is it because they appreciate certain stimuli – or certain dishes – more that these stimuli were stored? In short, is the observed effect truly the consequence of the erasure variable introduced into the experimental setting?

These different possibilities cannot actually be distinguished for epistemological reasons stemming from the "psi paradox" (Rabeyron, 2023): this ambiguity regarding the "source of psi" appears to be intrinsic to the conditions of its emergence. These phenomena emerge from

an order of reality, the ontic order, within which spatio-temporal distinctions are not effective. They arise from an "informational magma" not organized by the coordinates of space, time, and ordinary causality, which renders the previous distinctions and questions obsolete. The synchrony induced by the solidarity of the elements constituting the ontic order thus manifests as co-incidences, synchronicities, which can be sporadically observed but whose causal origins cannot truly be determined. Nevertheless, despite all these difficulties, it seems relevant to reflect further on the direction of the effect obtained and its potential consequences at the theoretical and empirical levels.

From this perspective, we might consider the following idea: the decomposition of the ontic order into the epistemic order occurs independently of time (we have to distinguish an ontic time from our epistemic time), such that the distinction between "before" and "after" is irrelevant between the "moment" of the participants' choice and the moment of data erasure. There is a "whole" that breaks down depending on multiple parameters (experimenters, participants, documentation, etc.). But one can imagine that within all this "magma," the storing of data has a predominant influence (which would explain why we are able to reproduce this effect). From a psychological point of view, this might mean that *we tend to bring into existence what we like*. Concretely, we would have, in a discreet and unconscious way, at the ontic level, a tendency to make what we appreciate "exist". Conversely, one can imagine that we tend to "destroy" what we do not like. This makes perfect sense from the perspective of the homeostatic logics that govern psychic life: what we love corresponds to values that are themselves correlated with elements of reality that promote our internal balance (and thus our survival); conversely, what we hate corresponds to values that are themselves correlated with elements of reality that undermine our balance (and thus our survival). *We therefore tend, in a subtle way, to "steer" the creation of the world in a way that promotes our survival.*

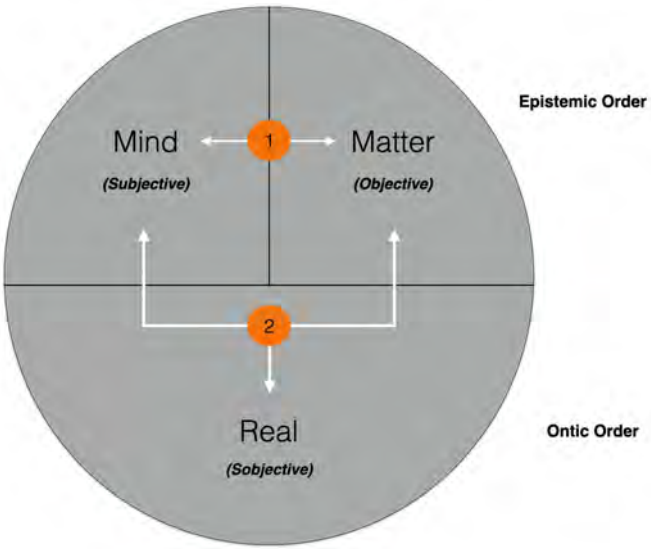
The two types of complementary-incompatible relationships

Another major theoretical contribution of this program of research lies in highlighting the relationship of complementarity and incompatibility between the subjective and objective measurements of the same stimulus. This relationship seems, more precisely, to be conceivable in terms of two distinct dimensions if we refer to the "Orpheus model" (see Figure 2; Rabeyron, 2023, 2026): a *horizontal* complementarity-incompatibility relationship between the subjective and the objective; a *vertical* complementarity-incompatibility relationship between the epistemic and the ontic. This dual relationship raises specific issues that likely have consequences for the emergence of psi interactions. The task would then be to determine how such a formal

conceptual framework leads to experimental hypotheses that extend the approach of Markus Maier and his team. We could also consider the implications of this model for understanding other phenomena through the logic of "oscillations"[3] between subjective experience and objective measurement, between epistemic experience and its ontic origin (for example, we could consider psychosomatic phenomena in which "oscillations" between the psyche and the body are sometimes observed clinically[4]).

Figure 2

Horizontal and vertical relationships in the Orpheus Model



3 Moritz Dechamps (2019) was able to demonstrate in his PhD "oscillations" in the evolution of psi effects as the data accumulated. These oscillations can be interpreted as the consequence of these relationships between the subjective epistemic order and the objective epistemic order.

4 It is somewhat like applying pressure to a water-filled mattress: this automatically alters the distribution of water within the mattress. The mattress represents a "totality" in which all elements are interconnected, producing automatic feedback loops. The same logic seems to apply in any experiment aimed at objectifying psi effects.

Relationship Between Acausal Entanglements and Causal Correlations

Another point that seems crucial from a theoretical standpoint concerns the nature of the relationship between causal and acausal relationships. I have proposed the hypothesis that *acausal relationships develop from causal relationships because the latter serve as the ground from which the former emerge* (Rabeyron, 2023, 2026). Acausal relationships are, in a sense, "folded" or "encrypted" within causal relationships. This would explain why maintaining a confound helps sustain the effect: it is not merely, as Markus Maier and his team suppose, that the confound protects the acausal correlation by rendering the source of the effect undecidable (thus avoiding the NT axiom). This causal relationship serves as the fuel upon which the acausal effect relies to emerge. It is thus a matter of *blending the copper of ordinary causality with the gold of psi*, for the latter cannot exist sustainably within the epistemic order without this causal underpinning.[5] It always ends up "evaporating," particularly as the degree of observation and measurement increases (Rabeyron, 2023, 2026).

This hypothesis helps explain why traditional parapsychology experiments ultimately fail. By rigorously separating the subject from the object, such protocols eliminate causal relationships, so that the effect persists for a time before eventually disappearing, as it no longer has the necessary ground for its emergence. This is precisely what is highlighted in the first experiment by Maier et al. (2026) published in this issue, which gives rise to the observed "*effect and decline*." But this effect would be difficult to observe directly because, as mentioned earlier, *acausal relationships are most often masked by causal relationships*. For example, in psychotherapy, "physical" exchanges (auditory, visual, etc.) between the clinician and the patient mask the "underlying" relationships that might emerge between them on the ontic level (Rabeyron, 2020). Such relationships only become visible when they are separated, provided that a sufficient therapeutic relationship has been established beforehand (an "organizational enclosure" according to the MPI). At the experimental level, the challenge therefore lies in maintaining causal relationships while simultaneously bringing to light the acausal relationships underlying them. This is precisely what Markus Maier and his team appear to have succeeded in doing: the confound maintains a form of causal relationship within the experimental setting, which could explain why the psi effect remains stable even when the experiment is replicated.

5 Another metaphor drawn from minerals comes from the relationship between corundum and chromium: a few atoms of chromium transform corundum (colorless) into ruby (bright red). Chromium is, in a sense, "encoded" within the crystal lattice of corundum and enables its transformation. In this regard, one might ask to what extent causal relationships also require acausal relationships to develop, as if the latter were their original, hidden point of emergence and contact with reality.

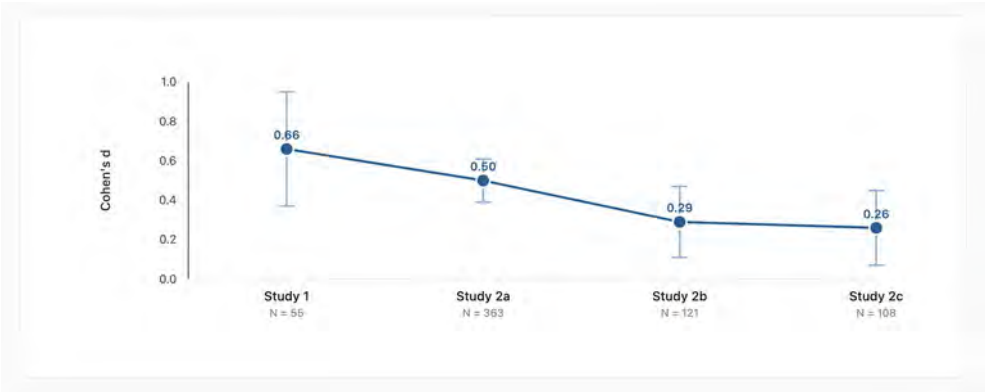
The Principle of Elusiveness and the Decline Effect

Nevertheless, as with any protocol aimed at objectifying the existence of psi, this research program still risks meeting the same grim fate as previous parapsychological paradigms. The same structural problem encountered in correlation matrix experiments is indeed likely to arise: at a certain level of analysis, the experiment ultimately produces a statistical prediction, specifically in binary form (participants' results with erasure (0) vs. participants' results without erasure (1)). This binary distinction constitutes a causal relationship between a prediction and a result. The same was true in the correlation matrix experiments: the psi effect could shift within the matrix, but one could still predict that the effects would be more pronounced in the matrix with participants compared to the matrix without participants. No matter how sophisticated the attempts to circumvent elusiveness may be, the experimental conclusion always takes the form of an empirically testable assertion: condition *X* will exhibit an effect, condition *Y* will not, which reintroduces a causal-predictive structure.

These theoretical perspectives suggest that a decline effect should be observed in the data collected by Markus Maier and his team: as the protocol is replicated, the causal evidence for the effect increases, which consequently would lead to a decrease in the psi effect. It turns out that the protocol involving gustative evaluation does indeed show such a decline effect: Study 1 ($d = 0.66$); Study 2a ($d = 0.50$); Study 2b ($d = 0.29$); Study 2c ($d = 0.26$; see Figure 3). The trend in effect sizes across the four studies shows a steady decline that does not appear to be due to chance. Based on the chronological order of the four studies, a permutation test yields $p = 1/4! = 1/24 \approx 0.042$ (one-tailed). A meta-regression weighted by the inverse of the variance of d (fixed-effects model) also reveals a significant linear decrease: $b = -0.138$ (95% CI: [-0.231; -0.046]; $z = -2.92$; $p = 0.0018$, one-tailed). This empirical signature is precisely what is predicted by the NT axiom and the idea that the acausal effect fades as the causal documentation of the protocol grows. This trend in effect sizes might lead us to formulate the following prediction: if the experiment were repeated identically, the effect sizes would continue to decrease (or even reverse) until they become insignificant.

Figure 3

Evolution of Effect Sizes in the Taste Evaluation Experiment



I therefore fear that the methods for circumventing the principle of elusiveness implemented by Markus Maier and his team could, in the final analysis, only delay the inevitable. The confound-erasure paradigm amounts to making a prediction about elusiveness itself, which transforms the decline effect into a relationship that can be predicted and anticipated in a stable manner. *Predicting elusiveness amounts to distorting elusiveness, for elusiveness cannot be predicted because it is necessarily elusive*. Psi is fundamentally a new emergence, a pure creation arising from the ontic order that "projects" itself into the epistemic order according to a specific "psi function" (Rabeyron, in press). But novelty can only be new once, just as one cannot reproduce an artistic work identically while producing the same effect of novelty. Psi is a singularity whose repetition is, for ontological reasons, incompatible with the very act of reproducing it, in the same way that a joke can only make the same audience laugh once. In this regard, it is somewhat as if Markus Maier and his team were attempting, by varying the stimuli (colors, dishes), to tell the same joke in different forms in the hope of continuing to make the audience laugh. This strategy allows the comic effect to be maintained temporarily, but the audience will eventually grow tired of it…

The Principle of Elusiveness: A Difference in Nature or in Degree?

Another theoretical point I would like to bring up for discussion concerns the formulation of the NT axiom itself. The term "non-transmission" could be misleading, as it suggests that no transmission of information is possible from the ontic order to the epistemic order. Psi does not allow

for "transmission" or "extraction" according to causal logic, and therefore psi cannot be used as a stable signal as one would with a causal and linear signal. However, psi can nevertheless allow for the transmission of information, thus falling into the category of "pseudo-signal" (Lucadou et al., 2007). From this perspective, the relationship between measurement and the emergence of psi is not one of fundamental incompatibility, but rather a *relationship that operates according to a difference in degree*. Information extraction via pseudo-signals is possible, but their stable and repeated transmission seems impossible. Nevertheless, the research program developed by Markus Maier raises the possibility of stable information extraction from the ontic order. The key question now is to determine whether this extraction can be sufficiently stabilized by circumvention measures (confound + erasure), and if so, whether they can be used for other tasks and under what conditions.

Observation, Recording, Documentation

Another methodological point deserves further clarification: the distinction between *storing* (the fact that data is stored in a persistent file), *observation* (the fact that a human agent becomes aware of the data), and *documentation* (the fact that the data is indexed and accessible to other observers). These three dimensions are not equivalent from a theoretical standpoint, and it would be interesting to test whether the erasure effect depends on one or the other. One can imagine, for example, data that is recorded but placed in a digital "vault" that would make its observation physically impossible. Conversely, would data erased after being briefly observed prior to its destruction retain a residual effect? Is observation cumulative: would a stronger effect emerge if observed by several successive experimenters?

Markus Maier and his team could attempt to include these parameters to determine their effects on the results. Nevertheless, I emphasize once again the consequences of the psi paradox: an experimental setting studying psi does not, *a priori*, allow – from an epistemological standpoint for reasons already mentioned – for distinguishing the respective contributions of recording, documentation, and observation. It would therefore be conceivable to introduce these variables to observe their effects, but I fear that these experimental variations will not lead to definitive conclusions. However, being able to demonstrate this problem experimentally will still be useful and will definitively rule out the possibility of experimentally distinguishing the influence of these different variables.

A Posteriori Permutation and the Temporality of the Ontic Order

Another theoretical point that Markus Maier and his team could elaborate on concerns the epistemic function of *a posteriori* permutation. In this regard, I would first like to recall the steps of the experimental setting:

1. A confound is deliberately maintained to prevent determining whether the effect is causal (confound) or acausal (psi): this produces a form of entanglement between the causal and acausal effects that prevents them from being "untangled";
2. The erasure of part of the data allows for differentiating the effect of the storing, observation, and documentation.
3. We (partially) "de-entangle"[6] the causal (confound) and acausal (psi) effects after the experiment has been conducted through permutation.

Maier and colleagues therefore assume that the introduction of confound "protects" the acausal correlation but that it can then be revealed by this permutation test. This strategy implicitly rests on the idea that there is no linear temporality at the level of the ontic order, insofar as the subsequent erasure of the data has a "retro-causal" effect on them. But if time is not a fundamental parameter of the ontic order, performing a permutation after the fact amounts to the same thing: it leads to conducting a test that documents the psi effect according to a causal relationship, which, in theory, should objectify the effect and make it disappear. For the reasons mentioned earlier, I fear that this will eventually happen, but let us imagine that it does not. This would mean that the temporal lag between the effect and its detection, combined with the erasure and the introduction of a confound, would allow the effect to be stabilized and to avoid the principle of elusiveness.

Experimenter Effect, Intentionality, and the Source of Psi

In the discussion of the first article, Maier et al. (2026) address the influence of the experimenter effect. Their conclusion is that this effect, although theoretically possible, seems unlikely, given the operational complexity that it would entail for the experimenters to exert such an influence on the experimental setting. This conclusion, however, rests on reasoning that remains tied to a

6 I say "partially" because, in reality, it can be statistically demonstrated that the effect is not reducible to confound, but it is impossible to determine precisely when the effect is causal or acausal. This is an important point from a theoretical standpoint, showing that the demonstration of the psi effect is probabilistic and statistical.

causal representation of such an experimenter's influence. Yet, psi effects seem to obey finalistic logics and are, from this perspective, opposed to the logic of causal relations (causal: cause > effect; finalistic: effect > cause; Rabeyron, 2023, 2026). This means that the system on which the psi effect operates "adjusts in reverse" to intentionality. For example, in a poltergeist case, an unconscious idea results in a consequence in the subject's material environment: intention precedes cause. A metaphor will help clarify this idea: the psi effect operates like a chain of dominoes, but in the opposite direction of ordinary causality. It is enough to have the initial intention (to knock over the first domino) for the entire set of dominoes to eventually fall (the causal chain of material events resulting from the fall of the first domino).

Consequently, contrary to what Markus Maier and his team propose, it is not necessary for the experimenter to be aware of all the details of the setting in order to influence it. The experimenter will simply exert an influence "at the end of the chain," and the psi effect will "travel back up" the causal chain. The literature on the experimenter effect provides ample documentation of this phenomenon, starting with the experiments by West and Fisk (1953), who conducted a classic ESP task by varying only the experimenter analyzing the data as a parameter, which allowed them to show that one of them obtained significant results while the other did not. Wiseman and Schlitz (1997, 1999; Schlitz et al., 2006) similarly obtained different results with an identical protocol.

However, it seems to me that the very question of the source's location (participants vs. experimenters) stems from a line of reasoning that remains captive to the spatio-temporal coordinates governing the epistemic order. In the ontic order, there is no separation between distinct agents: there is a *communion of wills*, in Schopenhauer's sense (1819/1969, 1851/1974), within an entangled system, where the experimenter, the participants, and the apparatus form an indivisible whole. Any parapsychological experiment therefore reduces this totality to separate elements that maintain relationships of complementarity and incompatibility. It is thus epistemologically impossible to construct a protocol fully independent of the conditions of its enunciation that would allow for separating the effect of the participants from the effect of the experimenters due to this logic of communion.

Stimulus and Pragmatic Information

As previously mentioned, Maier and his team do not explain the theoretical reasons that led them to choose the pictures of food dishes. This choice seems judicious, however. Color preference involves limited engagement: subjective variations are small, and the pragmatic information generated by each trial remains very modest. Taste preference, on the other hand, involves a

more intense hedonic investment, richer memory associations, and more pronounced individual preferences. Theoretically, the richer the task is in pragmatic information (the more important it is to the subject), the greater the potential magnitude of the effects should be. One could therefore consider other stimuli (faces, highly emotionally charged images, etc.), perhaps even more engaging, and see if this alters the observed effects, while taking all the epistemological precautions previously mentioned regarding the possibility of empirically distinguishing such variables.

Regarding future research strategy, however, I would propose to Maier and his team to replicate the taste protocol exactly before varying the stimuli. Otherwise, this would amount to introducing novelty into the system, which would certainly favor the production of a psi effect but would no longer allow us to determine whether the effect persists due to this introduction of novelty or due to the protocol itself. It is precisely this ambiguity that, in my view, is illustrated by the research program of certain researchers who manage to repeatedly produce significant effects but by varying at least one parameter between each replication. The protocol would therefore need to be *strictly* identical and reproducible, so that other research teams could claim to replicate it under the same conditions. Only independent and repeated experimental validation could then demonstrate that the effect is stable. But this brings us back to the principle of elusiveness: by proceeding in this way, the effect is objectified and "used" in a causal manner. The reader will thus have understood the paradoxical logic in which we find ourselves when it comes to objectifying psi effects – a situation that amounts to trying to jump over one's own shadow… The data obtained by Markus Maier are still too preliminary to determine whether he and his team have achieved such a feat.

Conclusion

The research program developed by Maier and his team constitutes one of the most original and promising paradigms in psi research. It proposes an experimentally testable approach aimed at avoiding the reproducibility issues encountered in this field. The approach to effects as perceptually induced correlations, as well as the gradual development of a strategy based on (1) identifying an effect associated with a confound, (2) storing only one part of the data, and (3) analyzing the data retrospectively to distinguish the psi effect, is particularly ingenious and leads to promising results. The theoretical comments offered in this article do not constitute a challenge to this program, but rather an outline of complementary perspectives for future research. I hope that this protocol will make it possible to stabilize the effects, but the elements presented in this article call for a certain degree of caution for epistemological reasons that have, until now, made it very difficult to reliably document psi effects.

Nevertheless, this research program already offers an original perspective that could truly become a new paradigm based on the recognition and consideration of the principle of elusiveness. Perhaps we can then hope, as Markus Maier and his team have begun to envision, that this paradigm will gradually come to fruition as research methods are refined. In this regard, the erasure-confound paradigm paves the way to elaborate the experimental parapsychology research program on renewed theoretical foundations. The future will tell whether this path will indeed make it possible to overcome the aporias of the Rhinian paradigm, or whether it, in turn, will run up against the paradoxical logic that has characterized the study of these phenomena for nearly a century.

References

Atmanspacher, H. (2003). Mind and matter as asymptotically disjoint, inequivalent representations with broken time-reversal symmetry. *BioSystems, 68*(1), 19–30.

Atmanspacher, H. (2020). The Pauli-Jung conjecture and its relatives: A formally augmented outline. *Open Philosophy, 3*(1), 527–549.

Atmanspacher, H., & Fach, W. (2013). A structural-phenomenological typology of mind-matter correlations. *Journal of Analytical Psychology, 58*(2), 219–244.

Atmanspacher, H., & Primas, H. (1996). The hidden side of Wolfgang Pauli: An eminent physicist's extraordinary encounter with depth psychology. *Journal of Consciousness Studies, 3*(2), 112–126.

Atmanspacher, H., & Rickles, D. (2022). *Dual-aspect monism and the deep structure of meanin*g. Routledge.

Atmanspacher, H., Römer, H., & Walach, H. (2002). Weak quantum theory: Complementarity and entanglement in physics and beyond. *Foundations of Physics, 32*(3), 379–406.

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology, 100*(3), 407–425.

Bem, D., Tressoldi, P. E., Rabeyron, T., & Duggan, M. (2016). Feeling the future: A meta-analysis of 90 experiments on the anomalous anticipation of random future events. *F1000Research, 4*, 1188.

Bohm, D. (1980). *Wholeness and the implicate order*. Routledge & Kegan Paul.

Cardeña, E. (2018). The experimental evidence for parapsychological phenomena: A review. *American Psychologist, 73*(5), 663–677.

Cardeña, E., Lynn, S. J., & Krippner, S. (Eds.). (2014). *Varieties of anomalous experience: Examining the scientific evidence* (2nd ed.). American Psychological Association.

Dechamps, M. C. (2019). *Mind-matter interactions and their reproducibility* [Doctoral dissertation, Ludwig-Maximilians-Universität München]. Fakultät für Psychologie und Pädagogik.

Dechamps, M.C., Maier, M.A., Pflitsch, M., & Duggan, M. (2021). Observer dependent confounds of quantum randomness: Effect stability and replicability. *Journal of Anomalous Experience and Cognition, 1*(1–2), 114–155.

Filk, T., & Römer, H. (2011). Generalized quantum theory: Overview and latest developments. *Axiomathes, 21*(2), 211–220.

Kennedy, J.E., & Taddonio, J.L. (1976). Experimenter effects in parapsychological research. *Journal of Parapsychology, 40*(1), 1–33.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–196). Cambridge University Press.

Lucadou, W., Römer, H., & Walach, H. (2007). Synchronistic phenomena as entanglement correlations in generalized quantum theory. *Journal of Consciousness Studies, 14*(4), 50–74.

Lucadou, W. (2015). *The correlation-matrix method (CMM): A new light upon the repeatability problem of parapsychology.* 58th Annual Convention of the Parapsychological Association & 39th SPR International Annual Conference, University of Greenwich, United Kingdom.

Maier, M.A., & Dechamps, M.C. (2018). Observer effects on quantum randomness: Testing micro-psychokinetic effects of smokers on addiction-related stimuli. *Journal of Scientific Exploration, 32*(2), 265–297.

Maier, M.A., & Dechamps, M.C. (2025). Macroscopic complementary relation between subjective observations and objective measurements of colors. *Journal of Anomalistics, 25*(1), 15–60.

Maier, M.A., Dechamps, M.C., & Rabeyron, T. (2022). Quantum measurement as pragmatic information transfer: Observer effects on (s)objective reality formation. *Journal of Anomalous Experience and Cognition, 2*(1), 16–48.

Maier, M.A., Vogel, A., & Dechamps, M.C. (2026). Macroscopic complementarity between subjective and objective food image assessments: A conceptual replication of the color-erasure effect. *Journal of Anomalistics*, *26*(1), 37–74. https://doi.org/10.23793/zfa.2026.037

Maier, M.A., Vogel, A., Storch, J., & Dechamps, M.C. (2026). Non-classical correlation between subjective and objective color observations: Change of effect as a function of its empirical documentation. *Journal of Anomalistics, 26*(1), 14–36. https://doi.org/10.23793/zfa.2026.014

May, E.C., Utts, J.M., & Spottiswoode, S.J.P. (1995). Decision augmentation theory: Toward a model of anomalous mental phenomena. *Journal of Parapsychology, 59*(3), 195–220.

Mossbridge, J., Tressoldi, P., & Utts, J. (2012). Predictive physiological anticipation preceding seemingly unpredictable stimuli: A meta-analysis. *Frontiers in Psychology, 3*, 390.

Pauli, W. (1994). The influence of archetypal ideas on the scientific theories of Kepler. In C.P. Enz & K. von Meyenn (Eds.), *Writings on physics and philosophy* (pp. 219–279). Springer. (Original work published 1954).

Plato (1991). *The republic*. Basic Books. (Original work published ca. 375 BCE).

Rabeyron, T. (2020). Why most research findings about psi are false: The replicability crisis, the psi paradox, and the myth of Sisyphus. *Frontiers in Psychology, 11*, Article 562992.

Rabeyron, T. (2023). *Codex Anomalia: De l'énigme du psi à la relation psyché-matièr*e. InterEdition.

Rabeyron, T. (2026). *The Psi Enigma: When anomalous experiences transcend the everyday world*. Routledge.

Rhine, J. B. (1934). *Extra-sensory perception*. Branden Press.

Roll, W. G. (1977). Poltergeists. In B. B. Wolman (Ed.), *Handbook of parapsychology* (pp. 382–413). Van Nostrand Reinhold.

Schlitz, M., Wiseman, R., Watt, C., & Radin, D. (2006). Of two minds: Skeptic-proponent collaboration within parapsychology. *British Journal of Psychology, 97*(3), 313–322.

Römer, H. (2023). Twenty years of generalized quantum theory. *Journal of Anomalistics, 23*, 145–154.

Schopenhauer, A. (1819). *The world as will and representation* (Ed. 1969). Dover.

Schopenhauer, A. (1851). *Essay on spirit seeing and everything connected therewith* (Ed. 1974). Clarendon Press.

Spinoza, B. (1677). *Ethics* (Ed. 1996). Penguin Classics.

Storm, L., Tressoldi, P. E., & Di Risio, L. (2010). Meta-analysis of free-response studies, 1992–2008: Assessing the noise reduction model in parapsychology. *Psychological Bulletin, 136*(4), 471–485.

Tierney, I., Watt, C., & Flores, A. (2018). Measuring organizational closure in the MPI/GQT/CMM context. *Journal of Parapsychology, 82*(2), 198–202.

Tierney, I. (2012). Lessons from a case study: An annotated narrative. In W. H. Kramer, E. Bauer, & G. H. Hövelmann (Eds.), *Perspectives of clinical parapsychology* (pp. 20–29). Stichting Het Johan Borgman Fonds.

Walach, H., & Römer, H. (2000). Complementarity is a useful concept for consciousness studies: A reminder. *Neuroendocrinology Letters, 21*(3), 221–232.

Walach, H., & Römer, H. (2011). Generalized entanglement: A nonreductive option for a phenomenologically dualist and ontologically monist view of consciousness. In H. Walach, S. Schmidt, & W. B. Jonas (Eds.), *Neuroscience, consciousness and spirituality* (Vol. 1, pp. 81–95). Springer.

Walach, H., & von Stillfried, N. (2011). Generalised quantum theory – Basic idea and general intuition: A background story and overview. *Axiomathes, 21*(2), 185–209.

Walach, H., von Lucadou, W., & Römer, H. (2014). Parapsychological phenomena as examples of generalized nonlocal correlations: A theoretical framework. *Journal of Scientific Exploration, 28*(4), 605–631.

Walach, H., Kirmse, K. A., Sedlmeier, P., Vogt, H., Hinterberger, T., & von Lucadou, W. (2022). Nailing jelly: The replication problem seems to be unsurmountable. Two failed replications of the matrix experiment. *Journal of Scientific Exploration, 35*(4), 788–828.

West, D. J., & Fisk, G. W. (1953). A dual ESP experiment with clock cards. *Journal of the Society for Psychical Research, 37*, 185–197.

Wiseman, R., & Schlitz, M. (1997). Experimenter effects and the remote detection of staring. *Journal of Parapsychology, 61*(3), 197–207.

Wiseman, R., & Schlitz, M. (1999). *Replication of the experimenter effect and the remote detection of staring*. Proceedings of the Parapsychological Association's 42nd Annual Convention, 147–153.

Zusammenfassung

Jenseits des Prinzips der Flüchtigkeit? Eine Analyse der Ergebnisse des Erasure-Confounds-Paradigmas

Dieser Kommentar analysiert die Ergebnisse zweier Versuchsprotokolle der Forschergruppe um Markus Maier, die sich auf das Erasure-Paradigma konzentrieren und zwei konzeptuelle Neuerungen kombinieren: die Verlagerung der Aufgabe von einer willentlichen Intention zu einer perzeptiven Bewertung sowie die gezielte Einführung eines Störfaktors (Confound) zur Aufrechterhaltung des Psi-Effekts – ein Ansatz, den wir als „Erasure-Confounds-Paradigma" bezeichnen. Nach Einordnung dieser Arbeiten in den Kontext der Aporien des Rhine'schen Paradigmas – insbesondere hinsichtlich des Psi-Paradoxes und des Prinzips der Elusivität – werden die Beiträge und Grenzen dieses neuen Forschungsparadigmas diskutiert. Die Analyse hebt den originellen und vielversprechenden Charakter der erzielten Ergebnisse hervor und situiert sie in einem breiteren theoretischen Rahmen zur Frage der Elusivitätseffekte. Mehrere Perspektiven werden zudem aufgezeigt, um zu bestimmen, inwiefern die implementierte Strategie eine nachhaltige Umgehung des Prinzips der Elusivität erlaubt.

Schlüsselbegriffe: Psi, Generalisierte Quantentheorie, Prinzip der Flüchtigkeit, Psi-Paradox, Erasure-Paradigma, Decline-Effekt

Journal of Anomalistics
Zeitschrift für Anomalistik

www.anomalistik.de www.igpp.de

Comments on

(1) M. A. Maier, A. Vogel, J. Storch, & M. C. Dechamps: ***Non-Classical Correlation Between Subjective and Objective Color Observations: Change of Effect as a Function of its Empirical Documentation***

(2) M. A. Maier, A. Vogel, & M. C. Dechamps: ***Macroscopic Complementarity Between Subjective and Objective Food Image Assessments: A Conceptual Replication of the Color-Erasure Effect***

In: ***Journal of Anomalistics*****, *26*(1) (2026), 14–36 and 37–74**

Macroscopic Complementary – Questions and a Conjecture

Ian Tierney[a]

(a) University of Edinburgh

Abstract – The remarkable results obtained in the experiments reported here either confirm the hypothesis that it possible to demonstrate macroscopic complementarity or they are excellent examples of e-psi, otherwise termed experimenter effect. The simplicity of the methodology and the clarity of the statistical analysis in the 'documentation' experiment do not leave much room for alternative explanations. An investigation of a conjecture which might separate these 'explanations' is given.

Introduction

The two papers under review, a demonstration of macroscopic complementarity using taste stimuli and the effect of documentation on a similar experiment using color stimuli, are the latest in a lengthy series by Markus Maier as first author and his various colleagues concerned with the proposition that volitional autonomous reality construction: "via conscious perception involves a process akin to measurement wherein pragmatic information is conveyed". This proposal is based firmly on Lucadou's 1995/2007 Model of Pragmatic Information (MPI), initially a system theoretical derivation but later incorporated into Generalized Quantum Theory (GQT: Atmanspacher et al., 2002; Römer, 2023, 2024) which introduced concepts from quantum theory, particularly entanglement and complementarity into discussions about psychological phenomena. Incidentally and extremely regretfully this reviewer does not speak German so cannot comment on Römer's (2024) detailed exposition in *Quanten, Komplementarität und Verschränkung in der Lebenswelt. Verallgemeinerte Quantentheorie*. As a result there may be errors in presenting Römer's latest thoughts on these matters.

The supposition that complementarity could be extended to human psychology and experience was first proposed by Nils Bohr (1928) and the history of the idea, plus the background to the MPI and GQT is given in detail in the introductions to the Maier et al. (2024), Maier & Dechamps (2025) and the present papers. Essentially Römer (2023, p. 148) has defined complementarity within the GQT as an "epistemologically highly significant restriction of simultaneous predictability. It is not possible simultaneously to assign or deny with certainty all of the possible properties (accidentals) of a substance without mutual restrictions." It could be argued that while the definition of "a substance" is unclear, for this definition to be relevant to psychology it would have to include cognitive events such as an autonomous judgements, volitions and meanings.

The first set of experiments reviewed here employing taste stimuli are attempts to conceptually replicate the very significant and impressive results of the Maier & Dechamps (2025) experiments involving color stimuli. They were designed to examine the GQT contention that subjective experiences and 'related' (*reviewer's term: see below*) objective physical measurements constitute macroscopic complementarity through acausal, non-local entanglement correlations. "Complementarity in this context means the higher the impact of the objective component within the pragmatic information (…) the lower the impact of the subjective-autonomous element will be and vice versa" (Maier et al., 2024, pp. 19–20). This must be read taking account of their term "passive registration" where they contend that: "Objective, deterministic reality, as described by the physical sciences, involves passive registration excluding subjective volition (…) if measurements are passive registrations and therefore constitute objective measures, any

reality constructing intentional phenomena are automatically excluded and cannot be empirically observed." As the authors further point out, as a consequence of this these phenomena may be claimed non-existent whereas this is not an ontic inexistence but rather a measurement dependent one caused by the act of passive registration.

The second paper being considered in this review has a different if tangential aim which is to provide evidence for a decline in results which support a hypothesis involving acausal, non-local entanglement correlations when no attempt is made to avoid or reduce the firm explicit causal inference or signal such results may contain. Such a decline is seen as the inevitable result of the Non-Transmission Axiom (NT axiom), an essential part of the GQT, which arises from the contention that causal tests, interpretations or signals destroy acausal effects over time. Maier et al. (ibid.) contend that "to provide evidence for a volitionally constructed reality, which arises from an autonomous intentional impulse, one must reduce the confirmation element within the performed measurements (…). Attempting to objectively document volitionally created realities undermines their subjective impact" (Maier et al., 2024, pp. 19–20). In this 'documentation' paper the methodology, analyses etc. of the 'colour' paper (Maier & Dechamps, 2025) are used to explicitly demonstrate the 'effect and decline' when the method used to avoid this effect in the earlier paper is not employed. This avoidance was achieved in both the 'color' and 'taste' experiments by the deliberate introduction of color/calorific subset confounds within the experimental manipulations before demonstrating through a permutation analysis that the results could not be due to advantageous, chance 'picking'.

(1) *Macroscopic Complementarity Between Subjective and Objective Food Image Assessments: A Conceptual Replication of the Color-Erasure Effect*

Briefly, using the same method as the Maier and Dechamps (2025) design participants are engaged in an online procedure where they are asked to rate pictures of prepared food dishes for tastiness. As in any on-line experiment there are inevitable concerns about participant compliance but the authors' participant self-verification process is probably the best that can be hoped for and the large number of participants hopefully dilutes any such inadvertent variation. As in the color studies, limited subsets of food images are randomly selected from a larger pool and fixedly assigned to the experimental conditions within each study, thereby introducing the same deliberate confound that reduces the internal validity of each individual study and attenuates the directness of the causal test. The participants are divided into two groups. For one group the calorific value and stimulus ID are recorded but this is not done for the second (erasure) group. The participants are unaware of this manipulation which constitutes the independent variable

in the study. This procedure is repeated three further times with different stimulus sets. In all four sub-experiments the difference between the two groups in terms of the tastiness measure very significantly support the H1 hypothesis of the authors that participants would rate food images as more appetizing in the non-erasure condition compared to the erasure condition. Lastly, an overall analysis addresses the stimulus-bias confound retrospectively and provides a definitive test of whether the observed effects can be attributed to the erasure manipulation itself rather than to coincidentally favorable stimulus assignments to conditions.

If one agrees with the authors' premise that the measures of taste in this attempted replication experiment can be considered as subjective and objective and therefore, following the logic of Bohr (1928) might be complementary, then the methodologies employed including participant self-regulation, the statistical analyses, the statement of limitations etc., provide *compelling* evidence for their contention that the obtained differences between mean grouped subjective judgements may be due to the presence or absence of objective information about the stimuli of which the participants are unaware.

The experimental design of the experiments is relatively simple and the online assessments look robust with few opportunities for error or manipulation once the self-verification process is accepted. Maier & Dechamps conclude their 2025 paper thus:

> Finally, our findings align with the tenets of dual-aspect monism (see Atmanspacher & Rickles, 2022), which postulate a unified reality where in mind and matter exist in a dual yet complementary relationship. It should be noted, however, that the results presented here do not provide a definitive test for this theory. The GQT, with its mathematical and system-theoretical formalization was initially proposed more than 20 years ago (Atmanspacher et al., 2002; Walach & Römer, 2000) to address some of these questions about reality. While the empirical evidence for the GQT is still preliminary, our data suggest a promising outlook for its future.

Questions that Arise from the Methodology

In the present Introduction the reviewer used the term 'related' objective measures of the subjective judgement. This term questions the method and logic by which the objective measures were chosen. In selecting an objective measure of a subjective event, how were the particular taste attributes (i.e. calorific values and stimulus I.D.) chosen? In the 2025 paper Maier & Dechamps suggest that because they considered a color experience "a rather non-normative perception process" it was an appropriate stimulus to use and that various attributes of color were appropriate measures of this. It is less clear in the replication attempt whether taste and the

subsequent relevant objectivity measures conform to that description, and on what grounds: e. g., could the taste experiments have equally well used weight, proportions of protein vs. carbohydrates, or origin of recipe as objectifiers? Furthermore it is unclear on what grounds the direction of difference was established in the hypothesis that "(…) participants would rate food images as more appetizing in the non-erasure condition compared to the erasure condition". Was that choice an arbitrary one? The authors do refer to the earlier paper by saying the hypotheses were: "based on the theoretical framework of the GQT proposing non-commutability of measurements derived from meaningfully related macroscopic psychophysical subsystems and based on the empirical precedent established by the color-erasure research" (Maier & Dechamps, 2025).

Does the presence or absence of *any* objective measure that could be construed as relevant to the subjective choice affect the subjective measure? If so, do they affect the subjective choices equally etc? It could be argued that to demonstrate the author's contentions any objectification or absence of same should influence the subjective judgement, so did the choice matter? Laboring this point about whether or not the choices were arbitrary ones made by the experimenters is relevant to the possibility of an e-psi or 'experimenter effect' interpretation of these results which is central to the discussion of the second paper (see below). This is pertinent in the light of the authors' comments that "in order to convince a skeptical community of the existence of macroscopic complementarity (…) etc. (Maier & Dechamps, 2025, p. 21)" and "that the main goal of the research presented here was to provide empirical evidence for macroscopic psychophysical complementary relationship (…) etc." (ibid., p. 17). These are clear autonomous volitional statements.

The reasons for the ingenious employment of an additional variable which was introduced to give an alternative explanation for the results are given in detail. This was first used and explained in detail in the 2025 color paper. At the suggestion of Römer (2024, p. 56) it was employed to circumvent the supposed effect of clear unequivocal causal inferences from the results. The experimenter's subsequent employment of a permutation analysis provided compelling evidence that these effects were not due to chance ("lucky") stimulus assignments. If the NT axiom does indeed affect results as hypothesized, this manoeuvre is very welcome and needs to be recommended.

However, a somewhat abstruse corollary to this observation depends on whether causality necessarily requires an observer to establish that relationship. In describing their ingenious technique to "attenuate the direct causality test" they are assuming that the audience for their methodology which "leaves room for the alternative explanation" is presumably someone other than themselves. That is, while acknowledging the possibility of an alternate (chance) explana-

tion for differences their writing indicates that they do not share its likelihood. This rather odd observation looks deep into "the deep structure of meaning" and requires more examination. It is possible that it is covered by Römer (2024) which is unfortunately a 'closed book' to this reviewer. However, in the present context, this too would be relevant if an e-psi (experimenter effect) interpretation of the results is considered.

Question Arising from the Conclusion

More generally though – and this is a difficult point to make without sounding superficially adversarial – do the authors believe the effect they have clearly proven in their experimental design applies outside their laboratory and the specific settings of this experiment? The BF_{10} values confirming the H1 are very high, so one might assume the principle, the effect of objectification or otherwise on subjective judgement, to have real substance. If so, why has it not affected many, many different events and outcomes outside their experimental setting, in the world at large? This argument is based most obviously on the familiar appeal to evolution. If indeed complementarity affects subjective volitions, expectations and judgements depending on relevant objectification, that principle should confer either advantage or disadvantage to some reliable degree which, in a world where billions of such volitions or judgements are made each day, should be evident in everyday life. One might consider lottery betting choices where an individual's number choice is selected by that individual (with several types of objectification: e. g. ticket, either paper or on-line) or left to a random 'lucky dip' process logged minimally by an anonymous computer (minimal objectification). This returns to the question raised above – is the degree of objectification evident in the complementarity effect? In betting should these variables affect success? If so, it is highly likely that such bias would have been evident and exploited given the prizes involved and the number of tickets sold. The question about generalizability of these findings is not a trivial one once the veracity of the reported results is accepted, which this reviewer does.

(2) *Non-Classical Correlation Between Subjective and Objective Color Observations: Change of Effect as a Function of its Empirical Documentation*

Very usefully, the experimenters in this series of experiments chose to test whether the results obtained in a simpler version of the color experiment, where the attempt to avoid an explicit and clear causal signal was *not* made, would result in a decline effect or more precisely an 'effect and decline' as described in the present Introduction. In this second experiment, in order to test the implications that follow from the Non-Transmission (NT) axiom, an integral part of

the GQT, they deliberately looked to demonstrate an 'effect and decline' pattern in the pattern of BF_{10} support for the hypothesized difference between the group subjective judgements about color, when for one group various objective measurements, hue and lightness, are made and recorded but erased for the other group. Erasure took place immediately after each participant completed their study. This was undertaken over a very large number of trials (6000+).

In passing, the timing and method for data erasure taking place requires further examination. In an early test of the MPI using spontaneous cases (Watt & Tierney, 2013) the computer program used to run the experiment, as in the present study, collected then erased the information. In retrospect it might have been a clearer manipulation for the key presses which obviously have to be functional at each participant's level not to be accepted by the collation program. The relevance of that point lies in when erased material 'existed', before or after it *potentially*, but not actually, comes within the experimenters' purview.

Predicted Pattern of Decline

It could be argued that within parapsychology the most reliable finding in the last 100 years has been the non-replication effect allied to the 'effect and decline' one. This has led to the identification of what Kennedy (2003) has labeled as "the capricious, actively evasive, unsustainable nature of psi." Since its inception the NT axiom has been virtually the only cogent and coherent explanation for the replication problem and reported 'effect and decline' pattern (decline effect) in experimental parapsychology results, once fraud, questionable research practices, and error are ruled out. As mentioned previously, Walter von Lucadou's MPI, including the associated NT axiom, developed from its early 1980s formulation as a systems-theoretical description, to its close association with GQT (Atmanspacher et al., 2002; Römer, 2023, pp. 146; in passing, this reviewer is puzzled by Atmanspacher's silence about the NT axiom in his major writings referencing the GQT since the 2014 Pauli-Jung Conjecture papers). However, demonstrating that the predicted decline effects can be isolated has proved extremely difficult, both practically and conceptually. Perhaps the most problematic element has been understanding why the decline effect, if the NT principle is axiomatic, should be a gradual 'effect and decline' in significant results rather than an abrupt cessation. In several comments in the present papers Maier et al. use the term 'gradual' or 'ultimately' e. g. "Lucadou's 'non transmission' axiom suggests that attempting to objectify subjective phenomena ultimately leads to their destruction". Why ultimately and not instantly?

The question arises: – when in any test of macroscopic complementarity, or indeed any test for psi effects, does the instrumental pragmatic information (attribution of 'cause') develop? It

cannot be solely and initially at the time of a further formal attempt at replication but should also occur during the initial testing period, depending on various factors. In this regard the wording of the prediction in the present paper is very welcome: "Given the experimental conditions outlined the erasure dependent likability effect should be found initially and later drastically decline." Logically, two further questions arise: how long does this decline last and why should it recover? Although it was not the case in the present experiment, when temporally contiguous experiments are reported by the same experimenter demonstrating decline across sub-experiments in different papers using similar methodologies etc., how, or why, does the effect of the NT axiom 'reset'? Does this depend on the addition of new naïve experimenters, or other factors? Broadly this raises the question: What is the reset condition between repeated 'successful' experiments by the same experimenter each of which demonstrate phenomena but also demonstrate 'effect and decline'? Is this dependent on MPI factors like Novelty or doubtful Confirmation?

Unusual Distribution in Figures 1 & 2

In the present paper the prediction is for an "effect and decline" pattern to occur both within and across studies viz: that strong Bayesian evidence for the original differential effect will be found during initial accumulative data collection, and fall off thereafter. The different 'stopping rules' used which could pick up changes indicating decline are clearly described in this paper, but Figures 1& 2 suggest that the strong *BF* (i.e. roughly > 5) doesn't start until approximately n=2220 (which by anyone's judgement is a great many trials!), and that prior to that point there were 3 'low' *BF* points (favoring an H0). It is not clear what one can infer from that distribution, although the permutation analysis convincingly indicates that it is not due to chance. Given the axiomatic nature of the NT restrictions, one *might* predict that the effects of complementarity seen as difference between subjective measures with and without the existence of objective measures would yield significant differences from the start, so that even at n=100 trials the effect would favor H1 (albeit to a slight, anecdotal, degree) and might only decline once the evidence for a causal signal in favor of the complementary effects was very (>10) strong.

Alternatively, one might assume that an experimenter starting a formal test of any hypothesis, while not starting from a point of neutral prediction about the system, would be uncertain. So, in an e-psi formulation (the description but not explanation that results are correlated with the experimenters' volitions and expectations) this might be reflected in a variable BF distribution much like the one observed? It is unclear what is happening in those first 2000+ tests to produce so much, quite large, variability in H confirmation, and, by implication, why so many

trials are required to establish a causal inference which, if subsequent decline is dependent upon it, begins abruptly at after c. 2940 trials when the BF_{10} reaches a very high 39.77 indicating very strong evidence in support of H1. The reviewer is very grateful to Markus Maier (*personal communication*) for the information that "with regard to the present study, we did not observe the BF continuously but started to analyze the data only after 2000 participants had already performed the experiment." This could be a coincidence but this conjunction certainly raises the possibility and gives some substance to an e-psi interpretation.

If one were to view this pattern from the point of view of feedback to the experimenters, which, as Maier and Dechamps (2025) have pointed out is the inevitable consequence of using Bayesian sequential analysis, that feedback *might* influence and change experimenter's beliefs and expectations so that one could (and this would be extremely useful) make different predictions about subsequent results in terms of experimenter effects. Is it fanciful to consider that, as the BF, or any other evidence in favor of a hypothesis involving anomalous events, accumulates to a *very* strong point the experimenter considers – "this cannot be sustained or the world would look a very different, strange, place if the effect inferred from this BF was as salient all the time"? If belief or expectation in the form of e-psi *do* play a role, then arguably the effect may collapse to a 'consensus' view of reality at this point. This type of experience and its possible consequences have been explored by the reviewer (Tierney, 2020).

The NT Axiom

Though there is considerable evidence in the literature that positive evidence for imputed psi effects fall off with time, the pattern of decline has been hard to establish. By relying almost exclusively on the theoretical element (NT axiom) to 'explain' non-replication it may have inhibited discussion about other possibilities for the observed variations in results. Maier et al. (2024, p. 50) have explored some of these possibilities subsumed under the term e-psi or experimenter effect, arguing that "more effort needs to be spent in future studies to minimize potential e-psi effects, for example, by blinding conditions and/or the involvement of uninformed data analysts." The discussion below considers this point in more detail with direct reference to the methodology of the present experiment. In a personal communication on this matter, Maier has made the very relevant comment that "*maybe the NT axiom is just a more abstract description of an underlying psychological belief-related phenomenon. In other words, the NT is not an alternative, but a different framing of the same mechanism*." One could extend this important comment by suggesting that complementarity within psychological systems may reflect a balance or otherwise between mutually exclusive expectations.

Furthermore, in that same dialogue Maier et al. have said (and I am extremely grateful for their agreement to quote them at length):

> Personally, we think that the NT axiom is not an axiom but naturally follows from the fact that acausal phenomena as proposed by the GQT cannot be robustly documented by a causal, objective approach. As we said in the manuscript, acausality is destroyed by causal testing approaches. It is like a symmetry breaking. And, it happens since macroscopic phenomena considered here are not limited by the Plank constant. Acausal macroscopic phenomena can only occur when the Planck constant does not hold (this can be assumed when not purely physical systems and their complementary relations are considered). Since psycho-physical relations transcend the physical realm by definition, the Planck constant can be ignored – but as a consequence these relations are not purely objective anymore (but still real) – when testing them objectively, they display a pattern of effect and decline, which means they are real and not objective simultaneously. Or in other words, the acausal complementarity when tested scientifically (with a causal testing approach) is projected into time. And the time signature of a real but not purely objective effect is "effect and decline." The realness of its nature can then only be proven indirectly, either by using the confound procedure or by arguing that the false-positive is an unlikely false positive. This is the strategy that we applied. We think that the NT axiom reflects the epistemic restriction when testing acausal macroscopic relations with causal testing approaches. It is not that the effect on the ontic level declines but rather that the effect documentation simulates the inexistence of an acausal effect under study. The effect is real but not objective as shown by the anomalous "effect and decline" structures that look like false-positives but are actually not.

Choice of Hypothetico-Deductive versus Various Inductive Methodologies of Enquiry

Comments made by Maier et al. (2024, p. 49) about the GQT and non-testability highlight the need to develop procedures which both illuminate the process involved in reality formation but which avoid or circumvent the obstacles: hence their experiments. This raises the question whether using theory driven approaches, particularly hypothetico-deductive ones, are the way forward? Rabeyron (2020) has invoked the myth of Sisyphus to characterize such endeavor. The possibility of e-psi and indeed the employment of a gamut of parapsychological ESP terms to 'explain' results inevitably introduces uncertainty about causation in psi experiments. Might this be avoided if slow, admittedly laborious, inductive approaches of various types, are employed. Or are causal explanations (signals) inevitable? Put another way, does describing circumstances surrounding anomalous events *necessarily* result in a causal signal rather than just neutral description? Maier et al. (2024) by employing small changes in the experimental design over a series of studies were, in effect, doing this, driven by observations of various kinds.

Because the present results *might* be interpreted as illustrating quite starkly the possible effects on outcomes of beliefs and expectations, this suggests that alternate modes of investigation are required?

Discussion

The results of this ingenious and impressive series of studies (and those in the present 'documentation' study in particular) are, arguably, the first experimental evidence that the NT axiom as an explanation of "effect and (drastic) decline" may have substance. This reviewer's attempts both to use this axiom to reduce unwanted spontaneous anomalous events via documentation (Watt & Tierney, 2013), to demonstrate its effects in a series of the matrices experiments (Flores, et al., 2018) and to find substantive persuasive examples of systematic decline effect in the results of many parapsychological experiments (Tierney, 2020) have all failed. In the latter endeavor reliable patterns of 'effect and decline' proved difficult to establish unequivocally. However, anecdotally and in contrast, the reviewer has witnessed a very dramatic decline (total cessation) of very unusual, almost unbelievable phenomena in circumstances that could be interpreted as supplementary documentation (Tierney, 2012), so, again, this present experiment is very welcome. It promises what experimental parapsychology has lacked for 100 years, a 'reliable' artifact to examine in detail.

Because the NT axiom, for which evidence has been limited, carries virtually all the burden of non-replicability (when fraud, experimenter effect and doubtful research practices are ruled out) it is an important concept in experimental parapsychology or anomalistics. The present results, which support this element in the GQT, may in turn encourage the use of related constructs. This view seems to shift the emphasis in heuristically useful models, particularly the MPI, towards language redolent of GQT/quantum physics rather than a psychology-based model emphasizing meaning/novelty and confirmation, which was its original derivation from the early 1980s. I hope I am not appearing impolite when, given the ubiquity of replication difficulties and reported position effects in psi research, and therefore the importance of the NT axiom, I suggest that the '(NT) tail might be wagging the (GQT) dog'.

Put another way, the inferences drawn from the very significant results in these papers tend to emphasize descriptions, explanations and hypotheses drawn from GQT and quantum physics as opposed to ones drawn from psychology and phenomenological psychology in particular, although it seems from Römer's (2023) synopsis of his 2024 book that this may be a simplistic distinction. Ideas drawn from phenomenological psychology are relevant because, uniquely, they rely on knowledge by acquaintance rather than knowledge by description (Russell, 1910),

so encompassing all the types of subjective experience to which individuals have 'direct' access. While this too might be interpreted as the effect of two types of measurement or observations on a single datum, hence implicating complementarity, arguably it involves two subjective measurements or observations of different data sets, and it may be the consequence of dissonance between these observations (Tierney, 2020), that underpins anomalous results.

The interpretation of results in both the 2025 color experiment and in the conceptual replication using taste experiments, hinge on whether the results reflect, as the experimenters claim, and set out to test, and subsequently find evidence for, systematic differences between subjective and objective measures due to macroscopic complementarity, *or*, due to the autonomous volitional intention of the experimenters obtain results congruent with their H1 by induced correlations: this is the e-psi or experimenter effect interpretation. To repeat the point, the e-psi description, driven as it is by 'single' internal states within the experimenter(s) differs from the complementarity explanation which involves two or more judgements, observations or measurements. Either interpretation (complementarity or e-psi) explanation would favor the Atmanspacher & Rickles (2022) interpretations by supposing that the results revealed either structural or induced correlations. Also, as mentioned previously, it is also possible that the two interpretations share common features, whereby an evolutionary progression from induced to structural correlations occurs.

Two Related but Separable Scenarios

In these two papers Maier et al. may have, very usefully, set the scene for an examination, *but not a test*, of two possible scenarios. The first scenario, the one they present, is that GQT and all that it implies, has heuristic power in understanding the origins and relevant variables in anomalous experience. Römer (2024) has insisted that the GQT is not a physical theory, but an "epistemology", which may emphasize its heuristic strength – structuring understanding in a way compatible with human attributes and attributions. Included in this are concepts like 'meaning' which in the form of pragmatic information may not be amenable to computation, number and statistics (Atmanspacher & Rickles, 2022, p. 196). Nevertheless, the possible replication of complementarity in the form suggested by these experiments would go a long way to substantiate this "epistemology".

Their 'documentation' experiment both supports their H1 but also raises the question of when this support occurred and whether indeed, as Maier has suggested (*personal communication*), it may be an effect of increasing statistical power rather than the result of formative observations by experimenters who propose, and have a vested interest in establishing, H1.

Consequently, the second scenario is that given certain conditions the nature of a DV being tested (and by implication the theory it arises from) is immaterial, it is the "autonomous intentional impulse" of the experimenters that prevails, at least initially, usually when a hypothetico-deductive approach is employed. This scenario is recognizably the experimenter effect or e-psi scenario discussed, and appealed to, for many years in the field. However, it has to be emphasized that e-psi is a description not an explanation, not least because it is likely that acausal interpretations are required. It may be that Maier et al. have, in the form of their 'empirical documentation' experiment provided a methodology which, if carefully incorporated into inductive inquiry may circumvent the problems inherent in both scenarios.

A Conjecture and 'Test' Using the Present Methodology

This second scenario prompts a conjecture. The history of parapsychological research for many years, but particularly in the last 60 years, has several examples, possibly half a dozen, where (1) well regarded and established researchers; (2) working in a small group that is initially 'organizationally closed' (Lucadou, 1995; Tierney et al., 2018) to a definable degree; (3) following sound theoretical and methodological principles; (4) testing what appears to the average observer as a novel but highly unlikely hypothesis; (5) initially achieve very impressive results supportive of their hypothesis; (6) these hypotheses are derived from theoretical backgrounds that are generally well regarded by the scientific community but which are given a novel 'twist' to accommodate/explain anomalous experience. Uniformly, when attempts to replicate these results are made by other groups or individual experimenters with a varied range of beliefs and expectation, the combined result, viewed in a meta-analysis, is equivocal or demonstrates very small effect sizes in their favor, reducing support to an anecdotal level. The conjecture is that each of these 6 elements are *necessary* to obtain the initial significant results supporting the novel hypothesis, but that increasingly more 'objective' inspection and criticism, ostensibly to eliminate implied error or fraud, has the effect of limiting these findings resulting in doubts about replication. The failure to reliably replicate the substantial results reported by Helmut Schmidt (animal PK), Chuck Honorton et al. (Ganzfeld), Ed May (distant viewing), Daryl Bem (presentiment) and now, possibly, Markus Maier (macroscopic complementarity) may be the consequence of the Orpheus model (Rabeyron, 2026), whereby closer, rigorous, exact and constraining investigation reduces the anomalous result or effect. Attempts at replication by sceptical and non-sceptical researchers attempting to pin down the causal logic behind such significant results, in effect looking too analytically at the results, doom the results to conform to consensus interpretations. The essence and tenor of this conjecture is summed up by saying that in the necessary circumstances the *ritual* of scientific inquiry is completed successfully and

the novel hypothesis is confirmed. The examples quoted may be the best illustrations that exist of the principles described in the MPI in its early form.

If autonomous groups of experimenters with various established views/expectations on the likelihood of replicating the present results of the 'documentation' experiment could attempt to repeat this experiment, the results, viewed inductively, might be instructive. Crucially, each group should choose when, during the accumulation of trials, they start the examination of data/results without conveying that choice point to other groups. Results should immediately be uploaded to a cloud site and each group must undertake not to share their choices and indeed the results of their experiments until a 'reveal' point when all the observations are examined. No hypotheses, other than the ones given by Maier et al., are required although necessarily the expectations of the different groups will differ, and clearly such an undertaking requires an administrative team who could be naïve to a number of matters, but 'an experimenter' is not required! This is not a perfect solution but one which is likely to yield many observations with varying degrees of successful replication or otherwise. As with all induction the Humean caveats are relevant but unavoidable and, arguably, are to be preferred to the competition of hypothetico-deductive inquiry. Hopefully, it would result in reliable descriptions but not explanations.

Conclusion

In their experiment to determine the effect of documentation on what they propose is macroscopic complementarity, Maier and colleagues *may* have found a way to illustrate the processes (causal and acausal) implicated in these anomalous phenomena.

Whether 'effect and decline' is due to axiomatic principles such as the NT axiom which, as replication efforts constrain choices and degrees of freedom within experiments, increase causal inference thus destroying acausal correlations, or, alternatively whether complementarity in psychological circumstances may be the 'epistomological' (in Römer's terms) description of the balance or imbalance of belief /expectation between individuals with different backgrounds and experience, remains to be seen. It would appear to be a dead end to suggest that these alternatives can be used to test (in the hypothetico-deductive sense) these conjectures. It is likely that establishing the general principal by using inductive logic is a more reliable method.

Either way this series of experiments and particularly the 'documentation' one advances understanding of the origin of anomalous experience. It increases the likelihood that it is possible to 'nail jelly' in the phrase employed by Walach et al. (2022) in that even patterns, and degrees, of failure to replicate results should be instructive Either macroscopic complementarity as envisaged in the GQT, described and demonstrated in these experiments, is substantiated

in replication by individuals other than the Maier group, *or*, complementarity in psychological systems is different in kind from that in envisaged for physical ones, being more akin to a balance or trade-off between meanings and expectations in individuals who are engaged in volitional autonomous reality construction. Replications of the Maier et al. 'documentation' experiment should be instructive here, depending on when (or if) 'effect and decline' results occur in relation to the initial point of observation/analysis of data/results. Presumably time will tell. Exciting times.

References

Atmanspacher, H., Römer, H., & Walach, H. (2002). Weak quantum theory: Complementarity and entanglement in physics and beyond. *Foundations of Physics, 32*, 379–406.

Atmanspacher, H., & Fuchs, C. (2014). *The Pauli-Jung Conjecture and its impact today*. Imprint Academic.

Atmanspacher, H., & Rickles, D. (2022). *Dual-aspect monism and the deep structure of meaning*. Routledge.

Bohr, N. (1928). The quantum postulate and the recent development of atomic theory. *Nature, 121*, 580–590.

Flores, A., Tierney, I., & Watt, C. (2018). *Where mind connects with matter: Replicating the correlation matrix method*. Presentation at the 61st Annual Convention of the Parapsychological Association, August 2–5. Institute of Noetic Sciences, Petaluma, California.

Kennedy, J. (2003). The capricious, actively evasive, unsustainable nature of psi: A summary and hypotheses. *Journal of Parapsychology, 67*, 53–74.

Lucadou, W.v. (1995). The Model of Pragmatic Information (MPI). *European Journal of Parapsychology, 11*, 58–75.

Lucadou, W.v., Römer, H., & Walach, H. (2007). Synchronistic phenomena as entanglement correlations in Generalized Quantum Theory. *Journal of Consciousness Studies, 14*(4), 50–74.

Lucadou, W.v. (2015). *The Correlation-Matrix Method (CMM) – a new light upon the repeatability problem of parapsychology*. Paper for the 58th Annual Convention of the Parapsychological Association and 39th SPR International Annual Conference, University of Greenwich, July 16–19.

Maier, M.A., Geissler, C., DeMattia, E., Vogel, A., & Dechamps M.C. (2024). Complementarity of subjective and objective realities. *Journal of Anomalistics, 24*, 15–54.

Maier, M.A., & Dechamps, M.C. (2025). Macroscopic complementary relation between subjective and objective measurement of colors. *Journal of Anomalistics, 25*(1), 15–60.

Rabeyron, T. (2020). Why most research findings about psi are false: The replicability crisis, the psi paradox and the Myth of Sisyphus. *Frontiers in Psychology, 11*, https://doi.org/10.3389/fpsyg.2020.562992

Rabeyron, T. (2026). *The psi enigma.* Routledge.

Römer, H. (2023). Twenty years of generalized quantum theory. *Journal of Anomalistics, 23*, 145–154.

Römer, H. (2024). *Quanten, Komplementarität und Verschränkung in der Lebenswelt: Verallgemeinerte Quantentheorie*. Lit.

Russell, B. (1910). Knowledge by acquaintance and knowledge by description. *Proceedings of the Aristotelian Society, 11*, 108–128.

Tierney, I. (2012). Lessons from a case study: an annotated narrative. In W. H. Kramer, E. Bauer & G. H. Hoevelmann (Eds.), *Perspectives of Clinical Parapsychology* (pp. 20–29). HJBF.

Tierney, I. (2020). Dissonance, complementarity and psi. *Journal of Parapsychology, 84*(2), 276–286.

Tierney, I., Watt, C., & Flores, A. (2018). Measuring organizational closure in the MPI/GQT/CMM context (Research Note). *Journal of Parapsychology, 82*(2), 198–202.

Walach, H., Kirmse, K., Sedlmeier, P., Vogt, H., Hinterberger, T., & Lucadou, W. v. (2022). Nailing jelly: the replication problem seems to be insurmountable: Two failed replications of the matrix experiment. *Journal of Scientific Exploration, 35*, 788–828.

Watt, C., & Tierney, I. (2013). A preliminary test of the Model of Pragmatic Information using cases of spontaneous anomalous experiences. *Journal of Consciousness Studies, 20*(11–12), 205–220.

Zusammenfassung

Makroskopische Komplementarität – Fragen und eine Vermutung

Die bemerkenswerten Ergebnisse der hier beschriebenen Experimente bestätigen entweder die Hypothese, dass es möglich ist, makroskopische Komplementarität nachzuweisen, oder sie sind hervorragende Beispiele für e-psi, auch als Experimentator-Effekt bezeichnet. Die Einfachheit der Methodik und die Klarheit der statistischen Analyse im „Dokumentations"-Experiment lassen wenig Raum für alternative Erklärungen. Es wird die Überprüfung einer Vermutung vorgestellt, die diese „Erklärungen" möglicherweise voneinander trennt.

Zeitschrift für Anomalistik
Journal of Anomalistics

www.anomalistik.de www.igpp.de

Comments on

(1) M.A. Maier, A. Vogel, J. Storch, & *M. C. Dechamps: Non-Classical Correlation Between Subjective and Objective Color Observations: Change of Effect as a Function of its Empirical Documentation*

(2) M.A. Maier, A. Vogel, & M. C. Dechamps: *Macroscopic Complementarity Between Subjective and Objective Food Image Assessments: A Conceptual Replication of the Color-Erasure Effect*

In: *Journal of Anomalistics*, *26*(1) (2026), 14–36 and 37–74

Chasing the Wind: Empirical Signatures of Generalized Entanglement

Harald Walach[a b]

(a) Next Society Institute, Kazimieras Simonavicius University, Vilnius, Lithuania
(b) Change Health Science Institute, Basel, Switzerland

hwalac@gmail.com

I applaud Markus Maier, Moritz Dechamps and their collaborators to this fine piece of thinking, resulting in even more impressive pieces of experimental documentation. I do this humbly:

A Bit of History

We have, over nearly two decades, striven to come up with a direct test of our Generalized Entanglement Hypothesis without success. The idea first occurred to me, actually in a conversation with Walter

von Lucadou in the train between Karlsruhe and Freiburg, when I discussed the strange findings of my clinical trial of classical homoeopathy in chronic headaches (Walach et al., 1997), around 1993 or so. I saw that the data signature of parapsychology and of research in homeopathy were rather similar: First experimental tests of whatever experimental paradigm was used found strong evidence for such anomalistic results. When following up on those, they tended to go away or switch channels, and the better controlled experiments were, the more likely such a decline was. My headache trial was in fact the third in a row of similar trials in homoeopathy research. The first (Brigo & Serpelloni, 1987) was extremely badly published, but had, according to Prof. Diener, the then head of the German association of neurologists, the best result in the whole migraine literature (implying: it must be wrong). The second (Whitmarsh et al., 1993) had a smaller effect, missing conventional significance, and my own trial was, as my fiercest critics grudgingly acknowledge, one of the methodologically best ever done in homeopathy and had the worst clinical outcome. After my conversation with Walter von Lucadou, it suddenly dawned on me: There is very likely a category of effects that are systematic, yet not causal (since clinical trials are, in a way, cause detectors). They are real, but cannot be manipulated in the same way as causal, for instance pharmaceutical systems, can. After a few more years of reading around in the quantum physics literature, studying Jung's synchronicity concept, and discussing with physicists I generated the idea that what was known in quantum theory as entanglement or non-local correlations was likely only a special case of a much broader, a generalized version of non-local correlations. My long-standing interest and study of Leibniz under the guidance of my late philosophy teacher Friedrich A. Uehlein hat helped here: Leibniz was the first who actually introduced such a concept under the term "preestablished harmony" (Walach et al., 2009). I approached Hartmann Römer, who then was the chair of theoretical physics in Freiburg with the idea; I had met him at a philosophy seminar invited by the philosopher Prof. Klaus Jacobi. When I detailed the idea that perhaps non-local correlations might also be operative in the macro-world under certain circumstances, he thoughtfully replied that this might be the case and that this should be formally stated. I challenged him to do that, and out came the formalism of Generalized Quantum Theory (GQT) (Atmanspacher et al., 2002). After that we spent quite some time and effort to "prove" that such generalized entanglement correlations exist. My hunch, at the time, was that we would have to find complicated models with many degrees of freedom. We tried Grinberg-Zylberbaum's (Grinberg-Zylberbaum, 1982) model of EEG-correlations across subjects. After initial positive results (Wackermann et al., 2003), and some really strong effects, that were never published (Naranjo, personal communication), the effects waned (Wackermann et al., 2004). We tried Walter von Lucadou's matrix PK experiment. After some strong initial results (Walach et al., 2020), a replication failed (Walach et al., 2021). We had indeed documented the No-Information-Transfer Axiom (NT axiom) in

real experimental life that we had previously confined to paper (Lucadou et al., 2007). This series of experiences together with others, not detailed here, convinced me that a) macroscopic entanglement correlations are likely real – under certain circumstances, and that my original idea was viable and useful and b) that it would likely not be taken up widely, because a direct experimental proof was impossible. And since our current scientific paradigm operates under the assumption that only experimentally validated and replicated results have to be taken seriously, we must await the dawn of another paradigmatic day.

For specialists: I know that physical entanglement correlations can be endlessly proven true, if one can operate the experimental set-up. But there is a decisive difference: Physical entanglement experiments are not experiments in the sense used in psychology, medicine or biology. They are documentations of empirical deviations from a theoretically expected behavior that are predicted or confined by Bell's theorem, i.e. measurements against a theoretical distribution. Since GQT is a much weaker formalism and does not have such strong quantitative framework notions as proper physical quantum theory, we cannot derive a theoretical prediction. And direct causal tests are ruled out by the NT axiom.

The NT Axiom and Time-Reversal Problems

The authors of these studies have taken the NT axiom seriously, probably because they saw it in action themselves previously (Dechamps & Maier, 2020; Maier & Dechamps, 2018; Maier et al., 2018). The ingenuity of these studies is the indirect approach. They bank on the reality of both generalized entanglement correlations and the prohibition of using them causally. Hartmann Römer has provided, in the appendix to our publication of the NT axiom (Lucadou et al., 2007), a formal proof, why entanglement correlations must not be used for causal signal transfer. If so used, they must break down. Every physicist knows this and the formal proof states, in a sense, the obvious. But why does it have to break down in the generalized case as well? Entanglement correlations are non-local, i. e. they appear as if signals could be transmitted faster than light. According to Special Relativity, the speed of light also sets the time frame of future and past, because nothing can travel at a greater speed than that of light. If we were able to use entanglement correlations causally, i. e. in a replicable and reliable manner, then we could break the arrow of time and could signal into the past. This leads to time-reversal paradoxes that have long been discussed in the philosophy of science (Fitzgerald, 1971): We could signal into the past and buy someone to kill our grandmother in which case we would not be here to do the buying. Therefore, nature does not seem to like such arrangements. A classical experiment, in which the control group is randomly created constitutes such a causal signal coding, at least

potentially. For the results of the first experiment could be used to know the group assignment of the second experiment, constituting a code that potentially allows signaling faster than the speed of light. That seems to be not a viable concept of Nature, and hence generalized entanglement correlations appear to have an inbuilt mechanism of self-destruction: As soon as they are misrepresented and misused as causal signals, they must break down. How quickly and how definitive such a breakdown happens seems to be dependent on the framework. If the controls, the blinding, the methodological rigor of an experiment are strong, the breakdown happens earlier.

But why do such systems work in practice? many ask. Homeopathy still works in clinical practice despite the fact it is very likely not causal. Synchronistic events happen all the time, and synchronistic systems such as the I Ching can be used to give meaningful answers to important questions. Parapsychological events, especially spontaneous ones, seem to happen despite the fact that they defy the current paradigm, and spontaneous macrokinetic events like spook and poltergeist phenomena have not died out. Single instances of them do seem to go away, if they are systematically documented, but as a group they continue appearing. The difference is this: None of these systems, supposedly operating on the basis of a generalized entanglement correlation, extracts causal signals out of the system. Homeopaths simply use homeopathic medicines. As soon as they start asking the question where, in the system, the supposedly healing information is, they are getting into trouble. Such systems work best if undoubted and used naively. But a scientific experiment is the opposite. It tries to extract a causal signal.

The Ingenuity of the Studies

This is, where the studies' ingenuity comes into play: Without compromising scientific rigor the authors have worked along the theoretical ideas of Generalized Quantum Theory and the predicted correlations and shown both: that such correlations appear to be operative. This can be seen in their high accumulated Bayes Factor of nearly 40 in the first study of the first publication and even higher in the second publication, which constitutes quite strong evidence. And the second prediction, that by repeatedly probing the system this would go away – an operationalization of the NT axiom –, was also borne out. The fact that they submitted also this specific distribution – a strong peak, followed by a steady decline – to a statistical test by generating a distinct distribution of this pattern by permuting the data 1000-fold is another ingenious step in analysis.

The second study does not test this hypothesis formally, but it also documents a decline of the effect size across the four studies from a high $d = 0.6$ to a smaller $d = 0.25$ with linear decreases from study to study.

I am convinced anyway. I would have to be modeled by a strong positive Bayesian prior and need no convincing. But what about a sceptic? Would he be convinced? Probably not. He might demand a replication of this study, even by a sceptic, to exclude a non-classical experimenter effect, whereby the consciousness and the expectation of the experimenters affect the system non-classically. Such a replication, indirectly, is the second set of studies in the second paper. The trick, why this second series does not continue the decline of the first set of experiments, is the novelty effect: instead of colors, food items are chosen. This, conceptually, constitutes a new set of experiments, and hence, sets out with a strong effect, documenting the effect of generalized entanglement correlation due to the incompatibility of objective measurement (in the second set of experiments: the caloric value) and subjective evaluation of tastiness.

In the first set of studies, the correlation rests in the meaningful complementarity between objective properties of presented colors and the subjective ratings and in the contrast between real and unreal pairs, in the case of deleted or unsaved properties. Since they are accumulated, each participant's data series constitutes a signature of such entanglement correlations, minute as it is. Knowing this, in a replication experiment the signature could be used to code a signal – stronger liking ratings for entangled versus unentangled pairings. So, the expectation according to the NT axiom would be that in a direct replication experiment the peak would go away. The authors avoided this effect in the second set of studies by using a different stimulus. Thus, they could set out afresh. Had they chosen to do more replications, seven, say, I am quite sure that the effect would have converged against a null-result. In the present set of four studies, they see the effect in each single study, albeit with declining effect size, and can confirm it in an overall analysis.

This, in my mind, constitutes preliminary experimental proof that generalized entanglement correlations exist and that the theoretical structure holds. This disallows causal misuse of the correlations. In future studies, new experimental paradigms will likely show again a strong initial effect that will decline with direct replications. It they are not too numerous, it might be possible to accumulate the effect, like in these studies. And across similar, but conceptually independent series of experiments the effect might be even more easily accumulated. Perhaps the empirical signature of the decline described by such quite similar experiments might help a good theoretician like Hartmann Römer to model an important parameter for the generalized quantum formalism, which will move the whole enterprise an important step forward. In that sense, I am looking forward to more ingenious studies from this creative and energetic group.

References

Atmanspacher, H., Römer, H., & Walach, H. (2002). Weak quantum theory: Complementarity and entanglement in physics and beyond. *Foundations of Physics, 32*, 379–406. https://doi.org/10.1023/A:1014809312397

Brigo, B., & Serpelloni, G. (1987). Le traitement homéopathique de la migraine: une étude de 60 cas, controlée en double aveugle (remède homéopathique vs. placebo). *Journal of the Liga Medicorum Homoeopatica Internationalis, 1*, 18–25.

Dechamps, M.C., & Maier, M.A. (2020). How smokers change their world and how the world responds: Testing the oscillatory nature of micro-psychokinetic observer effects on addiction-related stimuli. *Journal of Scientific Exploration, 33*, 406–434. https://doi.org/10.31275/2019/1513

Fitzgerald, P. (1971). Tachyons, backwards causation and freedom. *Boston Studies in the Philosophy of Science, 8*, 415–436.

Grinberg-Zylberbaum, J. (1982). Psychophysiological correlates of communication, gravitation and unity. *Psychoenergetics, 4*, 227–256.

Lucadou, W.v., Römer, H., & Walach, H. (2007). Synchronistic phenomena as entanglement correlations in Generalized Quantum Theory. *Journal of Consciousness Studies, 14*(4), 50–74.

Maier, M.A., & Dechamps, M.C. (2018). Observer effects on quantum randomness: Testing micro-psychokinetic effects of smokers on addiction-related stimuli. *Journal of Scientific Exploration, 32*, 265–297.

Maier, M.A., Dechamps, M.C., & Pflitsch, M. (2018). Intentional observer effects on quantum randomness: A Bayesian analysis reveals evidence against micro-psychokinesis [Original Research]. *Frontiers in Psychology, 9*(379). https://doi.org/10.3389/fpsyg.2018.00379

Wackermann, J., Naranjo, J.R.M., & Puetz, P. (2004). *Event-related correlations between brain electrical activities of separated human subjects: preliminary results of a replication study*. The 47th Annual Convention of the Parapsychological Association. Proceedings of presented papers.

Wackermann, J., Seiter, C., Keibel, H., & Walach, H. (2003). Correlations between brain electrical activities of two spatially separated human subjects. *Neuroscience Letters, 336*, 60–64.

Walach, H., Gaus, W., Haeusler, W., Lowes, T., Mussbach, D., Schamell, U., & Haag, G. (1997). Classical homoeopathic treatment of chronic headaches: A double-blind, randomized, placebo-controlled study. *Cephalalgia, 17*, 119–126.

Walach, H., Horan, M., Hinterberger, T., & von Ludacou, W. (2020). Evidence for anomalistic correlations between human behavior and a random event generator – Result of an independent replication of a micro-PK experiment. *Psychology of Consciousness: Theory, Research, and Practice, 7*(2), 173–188. https://doi.org/10.1037/cns0000199

Walach, H., Kirmse, K.A., Sedlmeier, P., Vogt, H., Hinterberger, T., & von Ludacou, W. (2021). Nailing jelly: The replication problem seems to be unsurmountable. Two failed replications of the matrix experiment. *Journal of Scientific Exploration, 35*(4), 788–828.

Walach, H., von Stillfried, N., & Römer, H. (2009). Preestablished harmony revisited: Generalised Entanglement is a modern version of preestablished harmony. *E-Logos: Electronic Journal for Philosophy, 7*. https://doi.org/http://nb.vse.cz/kfil/elogos/miscellany/walach09.pdf

Whitmarsh, T. E., Coleston, D. M., & Steiner, T. J. (1993). Homoeopathic prophylaxis of migraine. *Cephalalgia, 13* (Suppl.), 254.

Zeitschrift für Anomalistik
Journal of Anomalistics

www.anomalistik.de www.igpp.de

Authors' Response

Response to the Commentaries by Thomas Rabeyron, Harald Walach, and Ian Tierney on the Articles by Maier, Vogel, Storch, and Dechamps (*JAnom*, *26*[1], pp. 14–36; 2026) and Maier, Vogel, and Dechamps (*JAnom*, *26*[1], pp. 37–74; 2026)

Acausal Psychophysical Correlations in a Causally Closed World?

Evidence from (Un)Confounded Erasure Paradigms and the Limits of Physicalist Methodology

Markus A. Maier[a, *], Moritz C. Dechamps[a]

(a) Department of Psychology, Ludwig-Maximilians-Universität, München, Germany

[*] Corresponding author: Markus Maier
markus.maier@psy.lmu.de

Abstract – This response addresses commentaries by Thomas Rabeyron, Ian Tierney, and Harald Walach on two articles in this issue concerning macroscopic complementary relations between subjective and objective aspects of reality, framed within the Generalized Quantum Theory (GQT). The central concern shared by the commentators is whether the erasure-confound paradigm can robustly avoid the non-transmission (NT) axiom-induced decline that has affected earlier experimental paradigms in this research area. We argue that the NT axiom expresses an epistemic restriction arising whenever causal-deterministic documentation is applied to phenomena that contain an autonomous-subjective component. Under this reading, the erasure-confound paradigm achieves only a reduced-objective documentation of the effect at the single-study level, because the confound prevents an unambiguous causal interpretation, while the effect is established indirectly across studies through the exclusion of the alternative explanation (the "color selection

bias" confound). This design decouples evidence accumulation from the arrow of time and is proposed to circumvent the decline. We then put the decline prediction to an empirical test by reporting three previously unpublished color-erasure-with-confound studies (Studies 5–7), which together with the four studies of Maier and Dechamps (2025) form the complete record of seven studies in this paradigm, with no file drawer. The erasure effect appeared in the predicted direction in every study (sign test p = .008), and combined stratified permutation analyses across three independent 36-color sets ruled out the color-selection-bias confound as a joint explanation (Fisher p = .010; Edgington p = .018). An inverse-variance weighted meta-regression on the seven effect sizes returned a non-significant downward slope (b = −0.018, 95% CI [−0.043, 0.006], one-tailed p = .07), and the seventh study yielded an effect still clearly present (BF_{10} = 3,679.73, d = .30). The data are consistent with a stable, reduced-objectively documented acausal psychophysical correlation rather than the attrition predicted by the NT axiom under direct replication. Further specific points raised by individual commentators, such as the experimenter-psi interpretation and the choice of subjective and objective measurements, are addressed in shorter form.

Keywords: Generalized Quantum Theory, non-transmission axiom, NT axiom, macroscopic complementarity, erasure-confound paradigm, psychophysical correlations, decline effect, acausal correlations

We would like to begin by thanking the three commentators (Rabeyron, 2026; Tierney, 2026; Walach, 2026) for their valuable theoretical contributions to our empirical work presented in 2026. We greatly appreciate the time and effort they have invested in evaluating our research program and in initiating a discussion of its theoretical and future empirical implications. We regard supportive and critical arguments alike as the fruitful ground from which a better understanding of meaning-based psychophysical correlations might arise; correlations that we, following the Generalized Quantum Theory (GQT; Atmanspacher, Römer, & Walach, 2002; Römer, 2023; Walach & Römer, 2000), frame as acausal relationships between the subjective realm of reality (volition and experience) and the objective physical aspect of reality (material manifestations).

According to the GQT, the subjective and objective aspects of reality can form macroscopic complementary relations: When both aspects are meaningfully connected by an act of intentional observation, the corresponding psychophysical interaction is proposed to exhibit a number of specific features: (a) it is acausal in nature, meaning that corresponding subjective and objective aspects of reality are not causally related but non-locally correlated or entangled on macroscopic scales; (b) the dual nature of reality is epistemic and therefore observation-dependent, which implies that different forms of observation ranging from passive registration to fully intention-based perception yield different versions of duality, differing in the degree of macro-

scopic entanglement; (c) due to the macroscopically complementary framing of subjective–objective duality, observations of one aspect affect observations of the other, captured by what we have referred to as the incommutability conjecture; and (d) since these acausal psychophysical correlations form subjective–objective quasi-unified realities through entanglement-type correlations, these types of realities cannot be fully objectively documented when subjected to a standard scientific test. This last conjecture is formalized as the non-transmission (NT) axiom, which addresses the paradox involved in scientifically documenting acausal phenomena by means of a causality-testing experimental approach. Its axiomatic status follows from the fact that it cannot be derived mathematically from the GQT formalism. Rather, the NT axiom provides a necessary compensation for the ad hoc omission of Planck's constant to avoid contradictions and paradoxes within a causally closed objective world that scientific documentation needs as a precondition. In our view, however, the NT axiom can be derived exactly from this fundamental dilemma that an objective reality with its causal closure as a precondition for the objectification of phenomena is incompatible with any acausal and thus non-deterministic interference from the subjective realm. Or in other words, meaningful psychophysical correlations constitute sobjective realities and when they are projected epistemically by scientific documentation into the objective world (through causal experimental testing plus direct replication attempts), they gradually (since confirmation through direct replications follows a time signature) lose their sobjective status on the level of scientific confirmation.

The two articles addressed by the three commentators were designed specifically to test the incommutability conjecture and to validate the context derived from the GQT, under which the NT axiom should become active. One article (Maier, Vogel, & Dechamps, 2026) conceptually replicates earlier work (Maier & Dechamps, 2025) using what Rabeyron has aptly termed the "erasure-confound paradigm," here applied to food images with a focus on testing again the incommutability conjecture and the replicability of the erasure results under confound condition. The other article (Maier, Vogel, Storch, & Dechamps, 2026) provides a direct test of the impact of the NT axiom by employing a color-erasure-without-confound paradigm (the proposed context for NT to become active), conceptually replicating but extending the earlier color-erasure-with-confound research of Maier and Dechamps (2025).

Each commentary highlights important aspects of our work and offers ingenious interpretations, alternative theoretical perspectives, and pressing unresolved questions about the meaning of our results, about the theoretical background of the research, and about possible alternative interpretations of the data. Each author provides his individual perspective, but several common concerns also emerge, most prominently, predictions about the long-term reproducibility

of the effects obtained with the erasure-confound paradigm. We do not attempt to address every individual point raised in the commentaries here. Many of the arguments and theoretical interpretations made are highly creative and stand on their own; they reflect the authors' deep familiarity with the GQT, the Pauli–Jung conjecture, and related approaches to reality, and they offer fresh and innovative extensions of these frameworks. We share many of their views and appreciate their constructive engagement with our research program. Nor can we provide answers to all the open questions raised in the comments, partly because we simply do not know the answers and could only speculate, and partly because some questions require additional empirical data before definitive statements about the status of certain GQT-conjectures can be made.

In the present response, we focus on what we take to be the central theoretical concern shared by all three commentators, before turning to a complementary empirical contribution.

The Central Concern

The most central issue raised by Rabeyron (2026) and Walach (2026), and engaged with in a more nuanced way by Tierney (2026), is whether the NT axiom will eventually catch up with the erasure-confound paradigm. The question, in short, is this: Can the erasure-confound design robustly avoid the destructive consequences of the NT axiom, or will studies employing this design, like all previous experimental paradigms in this area, ultimately exhibit the same gradual decline that has plagued the matrix experiments (Walach et al., 2022), the earlier erasure paradigm (Maier et al., 2022), and a long line of related research within experimental parapsychology and anomalistics?

Behind this question lies a deeper one about what the NT axiom actually forbids, and about what "effect documentation" means within the NT axiomatic framework. Did we, in the erasure-confound studies, really confirm a true acausal effect, or did we merely document it indirectly by ruling out the only alternative explanation, leaving the effect itself sufficiently obscured that the NT axiom was not in fact activated? In what follows, we offer theoretical arguments for why the erasure-confound paradigm may indeed be capable of avoiding the activation of the NT axiom at the level of effect documentation. We then provide additional empirical data from three previously unpublished color-erasure-with-confound studies, which together with the four published color studies (Maier & Dechamps, 2025) allow us to put the decline predictions advanced by Rabeyron and Walach to a direct empirical test (Rabeyron, 2026; Walach, 2026).

Reflections about the NT Axiom

To frame our argument, it is useful to begin by reconsidering what the NT axiom is meant to capture. As Tierney notes in his comment (Tierney, 2026), we have suggested in personal communication (and the formulation is worth restating here) that the NT axiom is "not an axiom but naturally follows from the fact that acausal phenomena as proposed by the GQT cannot be robustly documented by a causal, objective approach." On this reading, the NT axiom is not an additional theoretical posit; rather, it expresses the epistemic restriction that arises whenever an attempt is made to test acausal macroscopic relations using a causal testing strategy. The following account of this restriction is necessarily speculative, but it is, we believe, broadly compatible with the framework developed in the Model of Pragmatic Information (MPI; von Lucadou, Römer, & Walach, 2007) and in the GQT (Atmanspacher et al., 2002; Römer, 2023; Walach & Römer, 2000).

In its original MPI formulation, the NT axiom captures a complementary relationship between two components of pragmatic information that arise in a system subjected to scientific documentation: novelty (the autonomy of the phenomenon in the system) and confirmation (the reliability of its documentation). The two components are inversely related: as confirmation/reliability increases through repeated empirical documentation, the novelty/autonomous impulse that originally generated the phenomenon must decrease. This is the formal source of the decline effects observed when acausal phenomena are subjected to direct replication.

More generally, the NT axiom protects the classical, causal-deterministic, and objective description of the world from documentations of phenomena that would contradict it. The classical-causal-deterministic and objective worldview is represented by physicalism, and it is precisely this view that the NT axiom protects with the implication that acausal phenomena of autonomous origin cannot be used for classical signal transfer. If they could, the nomological determinism that underpins, for instance, special relativity would be violated; hence the name non-transmission. The NT axiom thus ultimately protects the causal closure of a physicalist description of the world.

This causal closure is what enables the documentation of physical phenomena in objective form in the first place. The central recognized method of objectification is direct replication: the repeated identical empirical documentation of a phenomenon under largely identical causal conditions. One finds the same causally evoked phenomenon under the same conditions, a state of affairs that is possible only because everything is nomologically determined; otherwise, one would not consistently find the same experimentally produced phenomenon under the same conditions. This, in turn, generates causal closure: everything must be nomologically determined in order to be documentable in this way. The empirical criterion for the demonstration of objective reality therefore presupposes causal closure and determinism.

However, autonomous phenomena, i. e., phenomena that generate genuinely novel and unpredictable states, where "autonomy" is understood in the sense of Kane (2002) as the free origination of a decision without prior causal determination, contradict both nomological determinism and causal closure. In the case of meaningful psychophysical correlations in which an autonomous subjective experience, or an autonomous impulse in the sense of agency, interacts with the objective world, an inherent tension emerges. The autonomy located in (or proceeding from) the subjective realm conflicts with the structure of objective reality and its empirical form of documentation through direct replication. A causal description, or experimental operationalization, of such a psychophysical correlation must therefore itself exhibit a complementary relationship between autonomy and objectification. A high degree of autonomy cannot be objectified through direct replication, because the same thing simply cannot deterministically reappear under the same conditions, which is, after all, the defining feature of autonomy. Conversely, any attempt at objectification through direct replication would fail to capture the autonomy and would therefore declare it non-existent. A direct replication strategy applied to autonomous phenomena will, as the NT axiom predicts, produce a decline of the effect.

This is, in essence, the situation that Rabeyron (2026) describes when he writes that the very conditions under which psi interactions emerge, described as a "communion" or entanglement between the subject and its environment, are invalidated by the standard scientific framework, which presupposes a radical separation of subject and object. The incompatibility between a psychophysical correlation that contains both autonomous-subjective and deterministic-objective components leads during scientific testing to an initial effect documentation followed by a decline, a pattern that is phenomenologically indistinguishable from a false-positive finding that naturally regresses to the mean. This is precisely how the physicalist worldview is preserved: the phenomenon does appear initially, but it can later be declared a false positive and thus ignored. The NT axiom therefore formally guarantees causal closure at the epistemic level, while autonomous psychophysical interactions may continue to exist at the ontic level. In sum, when investigating specific forms of psychophysical correlations in the sense of unified interactions between autonomous-subjective and causally determined-objective aspects of reality, one is confronted with a fundamental obstacle: the incomplete objectifiability of this class of phenomena.

The Loophole: Reduced Objectification Does Not Trigger the NT Axiom

The central point of our argument lies precisely in this notion of incomplete objectifiability, or, in our preferred terminology, reduced-objective documentation. The NT axiom does not, as far as we can see, prohibit reduced or incomplete objective confirmation, so long as the

empirical data do not warrant an unambiguous interpretation in terms of a complete objective demonstration of the phenomenon. In the case of the erasure-confound studies, even a stable replication of the effect at the level of the individual study does not constitute a complete objective demonstration of an erasure effect, because the confound prevents any unambiguous causal interpretation at the single-study level. The subsequent statistical exclusion of the confound through permutation analyses does not alter this situation, because the requirements for an unambiguous, single-study demonstration of a clean erasure effect were never met to begin with. We therefore expect these effects to be robust over time, both within and across studies.

More generally, only methodologically clean individual studies combined with successful direct replication constitute the gold standard of objectification and thus the empirical basis for confirming an effect. Once these criteria are met only in a reduced form, as is the case in the erasure-confound paradigm by design, the complementary destructive effect from the objectification side does not, we propose, fully engage. This is, at least, our hope. Several additional considerations, however, might support it.

Subject–Object Communion, Autonomy, and Two Levels of Measurement

We share Rabeyron's (2026) analysis of psi phenomena and their behavior under scientific testing, and we extend it to the meaning-based psychophysical correlations investigated in the present studies. Rabeyron (2026) correctly notes that psi phenomena emerge from a subject–object communion or entanglement, and that scientific testing requires the separation of this communion, which is precisely the problem. We agree, and we take this to be exactly what the complementarity captured by the NT axiom expresses. It is important, however, to emphasize that the "subject" in this formulation must be understood as an autonomous subject. If subjectivity were construed merely as a mental mirror of the objective world (as it is, for instance, in analytic idealism or epiphenomenalist physicalism) the breaking of the communion would pose no problem at all, because subjectivity and objectivity would simply be two reflections of one and the same phenomenon. Only if subjectivity differs from objectivity in at least one essential respect – and that respect can only be autonomy, since autonomy is the one feature that is not causally determined – does the separation become problematic for scientific documentation. The mind–matter problem is a genuine problem only if autonomy is attributed to the mind. This point is, in our view, captured implicitly by the NT axiom and by the broader GQT and MPI frameworks.

With this in mind, two distinct measurement levels must be distinguished in the psychophysical correlations investigated in Maier and Dechamps (2025) and in the present studies (Maier, Vogel, Storch, & Dechamps, 2026; Maier, Vogel, & Dechamps, 2026). Here, "measure-

ment" is to be understood in the broadest sense as a transfer from unknown to manifest variables, encompassing both perception and the scientific documentation of an effect.

The first measurement level is the individual subjective perception of the color or food stimulus by the participant. When this perceptual act is permeated by individual motives and intentions, e. g. via evaluative judgments concerning how much one likes a color or how appetizing one finds a food image, it constitutes an induced correlation in the Pauli–Jung framework (see Atmanspacher & Rickles, 2022). The perceptual act then contains meaning, sense-making, and an autonomous contribution by the perceiving subject. If, by contrast, the perceiving person were to register the world only passively, without any autonomous willing, the result would be a structural correlation. To ensure the engagement of induced correlations and thereby the activation of the autonomy component required to make a macroscopic complementary relation visible at all, we deliberately focused on motive-supported perceptual processes (color likability and food deliciousness, i. e., preference judgments). Subjective brightness assessments or comparable perceptual judgments would, in our view, yield primarily structural correlations, which would commute normally, produce replicable effects, and exhibit no acausal interactions. The first prerequisite for finding any erasure effect is therefore the activation of an autonomous component in the psychophysical system under study, which is a precondition that has not, in our view, been sufficiently appreciated in previous attempts to investigate macroscopic complementary relations. In sum, the autonomous subjective component is connected with its objective counterpart by performing a motivated perception (level 1) of the outside world which establishes a to-be-studied meaningful psychophysical correlation.

The second measurement level is that of the scientific documentation of the effect: the demonstration of an acausal interaction between autonomous-subjective and causally determined-objective aspects of reality, that is, of the induced correlation itself. This is a collective level at which experimenters, and later observers of the results, are involved. At this level, documentation can proceed in either a fully objectifying way without confound, or in a reduced-objectifying way, as in the "with confound" case, where the erasure effect is veiled by the confound.

The Temporal Projection of Acausal Complementarity

What happens during the scientific documentation of such a psychophysical correlation? We agree here once again with Rabeyron's (2026) analysis: at the moment of documentation, the acausal interaction under investigation is projected into the classical, causal-deterministic, and spatiotemporally structured world. We formulated this point in earlier correspondence, and we are grateful to Tierney (2026) for citing the relevant passage verbatim in his comment.

Time, in particular, appears to us to be the central variable for understanding the consequences of the NT axiom. What occurs in documentation is that a timeless complementarity between classically incompatible opposites (autonomy and determined objectivity), which in timelessness pose no contradiction precisely because they do not causally interact, is projected onto the classical arrow of time. This projection produces causal necessitation not by temporal ordering alone but because scientific documentation operates under the precondition of causal closure. Its central method, direct replication, returns a stable result only if the documented events are nomologically determined, as we argued above. The arrow of time is the axis along which this method accumulates evidence, and the method's built-in assumption of determinism is what converts mere temporal order into the causal frame in which autonomy and objectifiability come into conflict. As a consequence, the level of description changes. What was previously timeless and therefore compatible[1] becomes uniformly temporally directed and incompatible, unless precautionary measures are taken.

In the color-erasure-without-confound study (Maier, Vogel, Storch, & Dechamps, 2026), in which no confound veils the causal structure of the erasure effect, the causal structure upon scientific documentation becomes manifest, and the incompatibility between the present autonomy and the determined objectivity comes into force. The original ontic compatibility due to timelessness is transferred into the four-dimensional spacetime with its clear uniform arrow of time so that both aspects, autonomy and objectifiable fixity, are forced onto the same direction. What the framework predicts robustly at this point is a negative claim: a clean causal test of an autonomous phenomenon cannot terminate in a stable, replicable, confirmed effect, because autonomy and objectification cannot both be satisfied under sustained documentation pressure. Instead, we are observing a pattern in the without-confound case consisting of an

1 The compatibility between autonomy and determined objectivity in a timeless realm can be illustrated through a two-arrows-of-time perspective. Along one time-direction, autonomous decisions appear as undetermined, freely chosen events that only establish a causal chain through autonomously produced links. Viewed backwards, the same chain appears fixed, since its outcomes, which are open from the forward perspective, are accessible as already-realized boundary conditions. The fixity at issue here is retrospective settledness, not nomological necessitation. A freely originated decision in the sense of Kane (2002) is, once made, equally settled in this retrospective sense without thereby having been necessitated by prior states and laws. The two descriptions are therefore not contradictory but perspective-dependent, in analogy to how two-state-vector formulations of quantum mechanics (e.g., Aharonov & Vaidman, 2008) allow both forward and backward boundary conditions to characterize the same quantum system. The forward and backward perspectives are not incompatible in themselves; they come into conflict only when both are collapsed onto a single forward arrow of time, as happens during scientific documentation. This is a speculative framing, but consistent with the broader GQT commitment to acausal ontic relations.

initial accumulation of evidence up to a maximum BF_{10} followed by a decline toward moderate evidence for H0. This pattern is consistent with this negative prediction, and is the one we interpret as autonomy being present early and then squeezed out as objective evidence accumulates. We emphasize, however, that this rise-then-decline signature is phenomenologically identical to a false positive, and we do not treat its specific shape as confirmatory.

It is precisely because a clean causal test yields a trajectory indistinguishable from a false positive that the discriminating empirical test must come from the erasure-confound paradigm, where the predicted pattern, directional consistency across independent replications without decline, is not one that a false-positive account would produce.

From the perspective of scientific documentation, the erasure-without-confound data therefore pose no epistemic threat to the causal closure of the physicalist worldview. Autonomy is, on this account, not objectively detectable under a clean causal test, and any trajectory it produces will be dismissible as a false positive. Under strict causal testing of an erasure effect, the evidence trajectory during scientific documentation is therefore one of effect-to-no-effect, corresponding to the arrow of time and to a failed replication, i. e., a "false positive".

The situation is fundamentally different in studies that employ the erasure-confound paradigm. At the level of the individual study, the reduced internal validity introduced by the confound establishes a reduced-objective evidence for the erasure effect, because the confound provides an alternative explanation, amounting, by standard scientific criteria, to a "no effect" at the single-study level. Through documentation, this psychophysical correlation is again projected onto the classical spacetime with its arrow of time. Studies and their repeated replications produce positive findings, and yet the effect itself cannot be convincingly demonstrated; each study remains in the status of "no effect". Only through the subsequent permutation analysis and similar procedures is the confound retrospectively and temporally subsequently excluded. The erasure effect is not directly demonstrated; rather, the only alternative explanation is excluded. The interpretation that there is a genuine erasure effect, i. e., that the effect is confirmed, is then empirical but only indirect, since only the alternative has been ruled out.

In this design, then, the temporal trajectory of effect production runs from no-effect to effect by exclusion, providing an overall reduced-objective demonstration of the macroscopically complementary and entangled phenomenon. From a physicalist perspective, the classical deterministic worldview and its causal closure remain intact, because each individual study has no probative power, and even the final analysis only excludes an alternative without directly showing the erasure effect. The erasure-confound method therefore achieves two things simultaneously: it decouples the direction of evidence accumulation from the direction of time,

thereby circumventing the false-positive problem; and it documents the effect in question only indirectly, through the exclusion of the alternative explanation. This, in our view, is the reason to expect such effects to remain robust in future replications.

We would add a speculative side note. We believe that spacetime as we know it, with its uniform arrow of time, emerges in its familiar form only through the collective-objectifying method of physicalism. If this is right, then time as it figures in scientific testing, that is, as a causal direction of evidence accumulation, is not a fixed feature of reality but a derivative of how the documentation practice is carried out. The order in which observations are made and aggregated should then itself be a manipulable variable, and decoupling test order from physical time, systematically and orthogonally, could be a powerful way to document autonomy-grounded psychophysical correlations and, more generally, psi effects (see also Schweizer, 2026).

Two further observations are worth making in connection with Rabeyron's (2026) analysis. First, his hypothesis that acausal relationships develop on the ground of causal relationships, with the latter serving as the substrate from which the former emerge, aligns closely with our argument. On Rabeyron's reading, the confound is not merely a veil that protects the acausal correlation from premature objectification; it is the very causal substrate, the "copper of ordinary causality" mixed with "the gold of psi," as he memorably puts it, on which the acausal effect can manifest. Our argument and Rabeyron's (2026) reinforce each other from two angles: from the NT axiom side, the confound prevents the direct causal test that would activate the decline; from the side of the phenomenon itself, the confound provides the causal ground from which the acausal correlation can stably emerge.

Second, Rabeyron (2026) raises the powerful objection that "predicting elusiveness amounts to distorting elusiveness," since elusiveness, if it is genuine, cannot be predicted while remaining elusive. We take this concern seriously. Our response is that the erasure-confound paradigm does not predict the acausal phenomenon directly. What is predicted is a directional difference between two conditions, with the source of that difference left ambiguous at the single-study level. What is then documented post hoc is not the acausal effect itself but the exclusion of the alternative causal explanation; the acausal effect is documented indirectly only by elimination, not by direct prediction-and-confirmation. Whether this distinction is sufficient to satisfy Rabeyron's (2026) elusiveness objection in the long run is, of course, an empirical question to which we now turn.

Empirical Evidence Against a Decline

We acknowledge that all of the foregoing is necessarily speculative, although we think it is internally consistent and grounded in the existing theoretical framework. As a final argument that

the NT axiom does not lead to a decline in the erasure-confound paradigm, we wish to bring empirical data to bear.

Rabeyron (2026) has conducted a decline analysis across the four published food-erasure-with-confound studies and has shown that these four studies exhibit a statistically significant declining trend in effect size (Study 1, $d = .66$; Study 2a, $d = .50$; Study 2b, $d = .29$; Study 2c, $d = .26$; weighted meta-regression $p = .0018$, one-tailed). The question is whether this trend is merely chance variation around a stable effect, or whether the food-erasure-with-confound effect is in fact in the process of declining steadily toward zero, as the NT axiom would predict for a paradigm that ultimately fails to circumvent the axiom. Walach, in his comment (Walach, 2026), makes a related and more specific prediction: at four studies the effect may still be visible, but at seven studies it would likely no longer be detectable.

We do not currently have additional food-erasure-with-confound studies available with which to test this prediction in the food domain. However, in the color-erasure-with-confound research, we have conducted, in addition to the four studies published by Maier and Dechamps (2025), three further studies that have not yet been published and which are described here for the first time. These are the only seven color-erasure-with-confound studies that have ever been carried out under our supervision, i. e. there is no file drawer. They therefore allow us to put both Rabeyron's and Walach's predictions to a direct empirical test in the color domain, which was the original paradigm in which the erasure-confound effect was discovered.

The remainder of this response presents these studies and the corresponding overall analyses across all seven color-erasure-with-confound experiments.

Three Additional Color-Erasure-with-Confound Studies (Studies 5–7)

In addition to the four color-erasure-with-confound studies published by Maier and Dechamps (2025; designated here Studies 1–4 of the color-erasure-with-confound series), three further studies were conducted under our supervision. These are the only three additional studies of this paradigm ever carried out, so that the seven studies presented below, four published and three previously unpublished, constitute the complete record of all color-erasure-with-confound experiments conducted under our supervision. No data have been collected but not reported, and no studies have been selectively excluded. We stress this point because the central claim we wish to advance, namely that the effect remains directionally consistent across all seven studies, would lose much of its force if a file drawer or selective reporting could not be ruled out. The complete data and analysis scripts for Studies 5–7 are available at OSF (https://osf.io/r7sqb).

Three design features distinguish Studies 5–7 from the original four-study series. First, they implemented a within-subjects rather than a between-subjects design in order to increase statistical power (Brysbaert, 2019). Second, they used 36 colors per study (18 randomly assigned to the non-erasure condition and 18 randomly assigned to the erasure condition), with all 36 colors presented to every participant. The increase from 6 colors (Studies 1–2) and 12 colors (Studies 3–4) to 36 colors further reduces the explanatory power of the "biased colors" alternative within each individual study. Third, all three new studies used new color sets drawn independently of those used in Studies 1–4 of Maier and Dechamps (2025), and Studies 6 and 7 in turn drew new color sets independently of each other, so that all three new studies tested fully independent stimulus material.

Beyond these modifications, the methodology of Studies 5–7 followed the food-erasure-with-confound studies reported in 2026 (Maier, Vogel, & Dechamps, 2026) closely, with the substitution of color stimuli (defined by the HSL parameters hue, saturation, and lightness) for the food images. The procedure, ethical approval, online data-collection platform, response scale, sequential Bayesian stopping rules, exclusion criteria (data-integrity item and color-blindness item), and the overall three-step analysis plan are therefore not described in detail again; we refer the reader to the methods section of the food paper (Maier, Vogel, & Dechamps, 2026) for the corresponding information. Only the study-specific details and results are reported below.

Study 5 (Squarra, 2025)

Study 5 was conducted as a Bachelor's thesis under the supervision of the first author (Squarra, 2025). Its aim was to conceptually replicate the overall color-erasure-with-confound effect documented by Maier and Dechamps (2025) using a within-subjects design and an extended stimulus set of 36 colors (18 per condition). The 36 colors were selected via a quantum random number generator (QRNG) prior to data collection. Hue varied randomly between 0° and 359°, lightness between 30% and 70%, and saturation was held constant at 100%. The 18 colors of the non-erasure condition were stored in the result file; the 18 colors of the erasure condition were stored only temporarily in a separate file during data collection (which was required for the program to display them) and were permanently deleted after the conclusion of data collection without prior inspection. Each participant rated all 36 colors on a visual analogue scale from "not at all" (0) to "very much" (100) in response to the prompt "How much do you like this color?". The minimum N to begin Bayesian sequential testing was 100, with an a priori stopping rule of $BF \geq 10$ in favor of H1 or H0, capped at approximately 250 participants for this initial sample.

A total of $N = 246$ participants (after exclusion of $n = 2$, one for self-reported color blindness and one for a failed data-integrity check) provided usable data. A one-sided Bayesian

paired-samples t-test (Cauchy[0, 0.1] prior) yielded BF_{10} = 1.34, indicating anecdotal evidence in support of H1 (d = .09; corresponding frequentist one-tailed p = .085). The mean likability rating was higher in the non-erasure condition (M = 50.82, SD = 12.99) than in the erasure condition (M = 50.08, SD = 12.17). The a priori stopping criterion of $BF \geq 10$ was not reached within the targeted sample.

A technical implementation error in Study 5, which was identified only after data collection had been completed, prevented the planned stratified permutation analysis and multilevel-model analysis from being carried out. Stimulus markers for the colors assigned to the erasure condition were not stored, so that the individual subjective ratings could no longer be aggregated by specific (although unknown) colors within the erasure condition. Without these aggregated color-level scores, the confound-exclusion step of the three-step analysis plan could not be performed at the single-study level. For this reason, Study 5 was not continued and has pilot status only. We nonetheless include it here for completeness, since excluding it would amount to selective reporting. Full methodological details and exploratory analyses (response time, age effects) are reported in Squarra (2025).

Study 6: Preregistered Replication

Study 6 was conducted as a preregistered replication of the within-subjects 36-color design tested in Study 5. The technical error of Study 5 was fixed, so that all 36 colors including those in the erasure condition were associated with stable, internally consistent stimulus identifiers that allowed the planned stratified permutation analysis to be performed at the single-study level. The preregistration is available at OSF (https://osf.io/mkv6x).

In brief, the procedure was a conceptual replication of Studies 1–4 of Maier and Dechamps (2025), with the following modifications: (a) a within-subjects design replaced the between-subjects design, with all 36 colors (18 per condition) presented to each participant; (b) the QRNG-generated colors varied randomly in hue (0°–359°) and lightness (30%–70%), with saturation held constant at 100%; (c) participation was restricted to PC, laptop, or tablet, with smartphone use prohibited (in line with prior observations that erasure effects appear weaker on smartphones); and (d) the Bayesian sequential stopping rule was asymmetric, terminating data collection when either $BF_{10} > 10$ in favor of H1 or $BF_{10} < 1/6$ in favor of H0 was reached, or upon reaching a maximum of N = 5,000 participants. The prior was narrow and centered around 0 (Cauchy[0, 0.1]) and the test was one-sided. The preregistered two-step analysis plan consisted of (1) a Bayesian dependent-samples t-test on subjective likability and (2), conditional on Step 1 yielding strong evidence for H1, a permutation analysis randomly reassigning all 36 colors to conditions across 10,000 permutations, with the criterion $p < .05$. A Bayes Factor

Design Analysis conducted at preregistration estimated an expected stopping point at $n = 683$ in the case of an effect and $n = 1{,}628$ in the case of a null finding. A total of 18 colors assigned to the non-erasure condition and another 18 colors assigned to the erasure condition were drawn randomly within the same parameter ranges and either stored to file or permanently erased without prior inspection (see OSF for color parameter details).

A total of 249 participants took part in the study. After exclusion of $n = 3$ participants (for self-reported color blindness or failed data-integrity check), the final analyzable sample consisted of $N = 246$ participants. The preregistered one-sided Bayesian paired-samples t-test yielded $BF_{10} = 15.19$, indicating strong evidence in support of H1 ($d = .18$; corresponding frequentist one-tailed $p = .002$). The mean likability rating was higher in the non-erasure condition ($M = 57.74$, $SD = 10.93$) than in the erasure condition ($M = 56.54$, $SD = 11.48$).

To address the biased-colors alternative at the level of this single study, a stratified permutation analysis was performed in which color assignments to conditions were randomly permuted 10,000 times, with the mean likability difference between conditions as the test statistic. The observed mean difference was 1.21 (stored – deleted); the permutation null distribution had a mean of 0.003 and a standard deviation of 2.64; the proportion of permutations yielding a difference at least as large as the observed difference was $p = .33$. The biased-colors alternative could thus not be ruled out at the single-study level.

The erasure effect of Study 6 was once again in the predicted direction, conceptually replicating the within-subjects pilot finding of Study 5 and the published cross-study finding of Maier and Dechamps (2025). At the single-study level, however, the stratified permutation analysis did not definitively rule out the confound, i.e., the biased-colors alternative. We therefore conducted Study 7 as a direct replication of Study 6 with an independently drawn new set of 36 colors.

Study 7: Direct Replication of Study 6

Study 7 was a direct replication of Study 6 with the sole modification of an independently drawn new set of 36 colors. The procedure, design, sample acquisition, response measures, sequential Bayesian stopping rule, single-study permutation analysis, and all other analyses were identical to those of Study 6. New sets of 18 colors assigned to the non-erasure condition and of 18 colors assigned to the erasure condition were drawn randomly within the same parameter ranges and either stored to file or permanently erased without prior inspection (see OSF for color parameter details). Study 7 was not preregistered, but the protocol, analysis plan, and stopping rule were identical to those of the preregistered Study 6.

A total of 249 participants took part in the study. After exclusion of n = 4 participants (for self-reported color blindness or failed data-integrity check), the final analyzable sample consisted of N = 245 participants. A one-sided Bayesian paired-samples t-test (Cauchy[0, 0.1] prior) yielded BF_{10} = 3,679.73, indicating extreme evidence in support of H1 (d = .30; corresponding frequentist one-tailed $p < .001$). The mean likability rating was higher in the non-erasure condition (M = 58.06, SD = 11.71) than in the erasure condition (M = 56.03, SD = 11.77). The BF trajectory crossed the BF_{10} = 10 threshold at n = 112 and continued to accumulate strong evidence throughout the remainder of data collection.

To address the biased-colors alternative at the level of this single study, the same stratified permutation analysis was applied. The observed mean difference was 2.03 (stored minus deleted), the permutation null distribution had a mean near zero and a standard deviation of 1.88, and the proportion of permutations yielding a difference at least as large as the observed difference was p = .14. The biased-colors alternative was therefore not ruled out at the single-study level.

Study 7 replicated the erasure effect documented in Studies 5 and 6, with the effect again in the predicted direction and with extreme Bayesian evidence. As in Study 6, the single-study permutation analysis did not exclude the biased-colors alternative on its own. This is the expected situation in this paradigm, where no individual study is designed to carry the full inferential burden, and the exclusion of the confound is established across studies rather than within any one of them, as the following overall analysis shows.

Overall Analyses Across All Seven Color-Erasure-with-Confound Studies

We next consider the evidence against the biased-colors alternative (the confound) across the full seven-study series. The three stratified permutation analyses available to us, which are the combined cross-study permutation of Studies 1 to 4 reported by Maier and Dechamps (2025) and the two single-study permutations of Studies 6 and 7 reported above, share an important methodological property. Each tests the biased-colors alternative on a set of exactly 36 colors. In Maier and Dechamps (2025), the 36 colors are distributed across four between-subjects studies (18 colors in the non-erasure condition and 18 colors in the erasure condition, divided across the four studies as 3+3+6+6 colors per condition). In Studies 6 and 7, the 36 colors of each study (18 per condition) are tested within a single within-subjects design. The three resulting permutation tests therefore arise from methodologically comparable analyses applied to three independently drawn 36-color sets, with the same null hypothesis (random color-to-condition assignment as an alternative explanation for the observed mean difference) and a comparable permutation null distribution structure.

The three corresponding permutation *p*-values were .005 for the Maier and Dechamps (2025) combined analysis, .33 for Study 6, and .14 for Study 7. None of the three excludes the alternative as decisively as a single clean study would, which is consistent with our view that the inferential weight in this paradigm rests at the cross-study level. Because the three 36-color sets were drawn independently via quantum random number generation on three separate occasions, the *p*-values are mutually independent and can be combined with standard methods. The Fisher (1932) combined-probabilities method yielded $\chi^2(6) = 16.75$, $p = .010$. The Edgington (1972) additive combination yielded $p = .018$. Both combinations fall below the conventional significance threshold and converge on the conclusion that across the three independent 36-color permutation analyses, the biased-colors alternative can be ruled out as a joint explanation of the observed pattern.

Two further observations support this conclusion. First, the direction of the erasure effect was hypothesis-consistent in every study, with the mean likability rating higher in the non-erasure than in the erasure condition in each of Studies 1 to 7 (see Table 1). Under the null hypothesis of no effect, the probability of seven directionally consistent results in seven independent studies is $0.5^7 \approx .008$ (one-tailed sign test). This test depends only on the direction of the observed differences and not on any of the within-study assumptions concerning normality, variance, or random color assignment, and it also includes Study 5, which could not include the permutation analyses due to the technical error described above. Second, the seven studies constitute the complete record of all color-erasure-with-confound experiments conducted under our supervision; there is no file drawer.

Taken together, the combined permutation analysis across three independent 36-color sets (Fisher $p = .010$; Edgington $p = .018$), the directional sign test across all seven studies ($p = .008$), and the absence of any file drawer, jointly support the conclusion that the consistent appearance of an effect in the predicted direction across seven studies, with independently drawn color sets, cannot be attributed to chance favorable color assignments. The biased-colors confound can be thereby ruled out at the level of the full series, even though no single study rules it out on its own.

Testing the Decline Hypothesis

The remaining empirical question is whether the seven-study sequence exhibits the systematic decline of effect size that Rabeyron (2026) and Walach (2026) predicted on the basis of the NT axiom. To test this prediction, we applied the same analytical approach Rabeyron used for the four food-erasure studies in his comment, an inverse-variance weighted meta-regression of the effect sizes on study order (1–7), using a fixed-effects model. One caveat applies to this whole

analysis and should be stated at the outset. In the erasure-confound paradigm the effect size of each study conflates the erasure effect with the intrinsic likability difference of that study's particular color draw (confound), so the seven d values are not clean estimates of the erasure effect. We nonetheless run the regression on them, because effect size is the quantity in which Rabeyron (2026) and Walach (2026) framed the decline prediction, and the appropriate way to address their argument is on its own terms.

The seven effect sizes entering the regression are: $d = .21$ (Study 1; Maier & Dechamps, 2025), $d = .50$ (Study 2), $d = .37$ (Study 3), $d = .14$ (Study 4), $d = .09$ (Study 5; Squarra, 2025), $d = .18$ (Study 6), and $d = .30$ (Study 7). The regression yielded a slope of $b = -0.018$ (95% CI [−0.043, 0.006]; $z = -1.46$; $p = .07$, one-tailed), indicating that the null hypothesis of no decline cannot be rejected at conventional significance. A complementary exact permutation test on all 5,040 possible orderings of the seven effect sizes yielded a one-tailed $p = .25$ for the observed slope, again well above conventional significance.

Two contextual comparisons help interpret these statistics. First, the magnitude of the estimated slope ($b = -0.018$ per study) is approximately one-eighth of the slope Rabeyron (2026) reports for the four food-erasure studies ($b = -0.138$ per study). Even taken at face value, the rate of decline in the color domain would be several times slower than in the food domain. Second, and independently of the magnitudes, the shape of the seven-study trajectory is not that of a steady decline. The largest effect was obtained in Study 2 ($d = .50$), followed by a gradual reduction across Studies 3 to 5 (.37, .14, .09), and a partial recovery in Studies 6 and 7 (.18, .30). The sequence of consecutive within-pair changes contains three decreases and three increases, again pointing to an irregular rather than a monotonic pattern.

These analyses jointly fail to support the prediction of a systematic decline across the seven color-erasure-with-confound studies. Walach's (2026) specific prediction that at seven studies the effect would no longer be detectable is not borne out by the data: the seventh and most recent study yielded an effect that was clearly present and directionally consistent. The most parsimonious description of the data is therefore a stable effect with irregular variation across studies, attributable to between-study sampling variability compounded by the study-to-study variability of the confound, rather than to a systematic decline.

Two qualitative observations remain worth making. First, the absence of a statistically reliable decline does not by itself refute the broader theoretical claim that the NT axiom must eventually catch up with the erasure-confound paradigm. It does, however, indicate that the catching-up has not yet occurred within seven studies, a meaningful empirical constraint, since seven is the number Walach proposed as a threshold beyond which detectability would be lost.

Second, the decline question and the existence question remain conceptually distinct. The biased-colors alternative is excluded by the combined permutation analyses and the decline question concerns the long-term behavior of an effect whose existence those analyses leave intact. Table 1 summarizes the key methodological features and the main effects of all seven color-erasure-with-confound studies.

Table 1

Overview of All Seven Color-Erasure-with-Confound Studies on Subjective Likability

Study	Source	Design	Colors per cond.	N	BF_{10}	d
1	Maier & Dechamps (2025)	between	3	945	22.28	.21
2	Maier & Dechamps (2025)	between	3	749	1.33×10^{8}	.50
3	Maier & Dechamps (2025)	between	6	562	603.16	.37
4	Maier & Dechamps (2025)	between	6	1,367	10.19	.14
5	Squarra (2025)	within	18	246	1.34[a]	.09
6	this response	within	18	246	15.19	.18
7	this response	within	18	245	3,679.73	.30

Note. Studies 1–3 used two-sided *BF*s (Cauchy[0, 0.1] prior); Study 4 used a one-sided *BF* (Cauchy[0, 0.2] prior, preregistered). Studies 5–7 used one-sided *BF*s (Cauchy[0, 0.1] prior); Study 6 was preregistered, Studies 5 and 7 were not preregistered.

[a] Pilot status; the a priori stopping criterion of $BF \geq 10$ was not reached. See description of Study 5 in the text.

Conclusion

The NT Axiom and the Erasure-Confound Paradigm: An Empirical Verdict

The empirical evidence presented in the second part of this response converges on a clear picture. Across all seven color-erasure-with-confound studies, the subjective likability of colors was higher when the corresponding objective color parameters were stored than when they were erased; this directional pattern held in every single study (sign test $p = .008$). The combined permutation analyses across three independent 36-color sets ruled out the biased-colors

alternative as a joint explanation (Fisher $p = .01$; Edgington $p = .02$). These two results are the confound-robust core of our case, since the sign test depends only on the direction of the differences and the permutation analyses isolate the erasure signal from the biased-colors draw. The decline question posed by Rabeyron (2026) and Walach (2026) was framed in terms of effect size, and we address it on those terms: an inverse-variance weighted meta-regression on the seven effect sizes returned a slope of $b = -0.018$ with a 95% confidence interval [−0.043, 0.006] that includes zero (one-tailed $p = .07$). The trend is slightly downward but does not reach significance, so the systematic decline predicted by the NT axiom, and specifically by Rabeyron (2026) and Walach (2026), is not statistically supported across the seven-study series. However, we attach limited weight to the magnitude of this slope in either direction, because in the erasure-confound paradigm the raw effect size and Bayes factor of each study conflate the erasure effect with the intrinsic likability difference of the particular color draw (=confound). The quantities that speak to the erasure effect net of the confound are the permutation *p*-values and the consistent direction, and both indicate a stable, present effect rather than a vanishing one. Taken together with the parallel evidence from the four food studies (Maier, Vogel, & Dechamps, 2026), the data support our theoretical argument in the first part of this response: the erasure-confound paradigm, by establishing only a reduced-objective documentation of the effect at the level of each individual study, does not appear to activate the NT axiom-driven decline that has plagued previous direct-replication paradigms in this research area.

One natural objection to this conclusion is that the seven color studies are not strictly identical in design, and that the absence of a decline may therefore be an artifact of repeated conceptual rather than direct replication. On this objection, the introduction of new color sets, the transition from between-subjects to within-subjects designs, the use of different stimulus counts per condition (3, 6, 12, or 18 colors), or even small differences in participant pools effectively "resets" the system across studies, prevents the accumulated documentation pressure that would activate the NT axiom, and thereby produces a misleadingly stable empirical pattern. We consider this a valid objection, and our reply is grounded in the structure of our own data. Studies 5, 6, and 7 share an identical design, a within-subjects 18-vs-18 color comparison with 36 randomly drawn colors per study, conducted on adult online samples of approximately 245 participants each ($N = 246$, 246, and 245, respectively). They are not conceptual replications but direct replications of one another. Under the design-variation objection, this three-study sub-sequence, in which the experimental design is held constant, should provide the cleanest test of the NT axiom-driven decline hypothesis and any genuine decline should be visible here.

What we observe in this sub-sequence is not the decline the NT axiom would predict. The effect appeared in the predicted direction in all three direct replications, and the seventh and

most recent study, far from showing the disappearance Walach (2026) anticipated by this point in the series, produced an effect that was still clearly present (BF_{10} well above the preregistered threshold) and directionally consistent. We also observe that the nominal effect sizes (d = .09, .18, .30) and *BF*s rise across the three direct replications. However, these magnitudes are confounded with the biased-colors draw of each study, as Study 7 makes plain: its extreme *BF* coincides with a single-study permutation p of .14, so the size of that evidence cannot be attributed to the erasure effect alone. What the three identically designed direct replications do establish is the absence of the predicted attrition under identical experimental conditions. Across the sub-sequence the effect did not weaken toward non-detectability but remained present and directionally consistent. The fluctuations across the full seven-study series are then best understood as ordinary between-study sampling variability, compounded by the study-to-study variability of the confound, rather than as NT axiom-driven decline. Walach's (2026) specific prediction that the effect would no longer be detectable by the seventh study is therefore not borne out, with the qualification that detectability in this paradigm is itself established jointly, through the combined permutation analysis and the directional sign test, rather than through any single study's effect size.

Experimenter Psi and the Source of the Effect

A second issue raised, in different forms, by Rabeyron and by Tierney concerns the so-called "experimenter psi" (e-psi) interpretation of our findings. Both commentators ask whether the observed effects might originate not in the participants but in the experimenters, whose expectations and intentions could non-classically influence the system. Tierney, in particular, develops this into one of two distinct interpretive scenarios for our results: either (a) the GQT-based reading we have proposed, in which the effect reflects a macroscopic non-classical correlation between subjective and objective measurements within an entangled participant–stimulus system, or (b) an e-psi reading, in which the effect is generated by the autonomous intentional impulse of the experimenters themselves. Tierney notes that hypothetico-deductive methods may not be able to discriminate between these scenarios, and he therefore suggests inductive approaches as a possible way forward.

We share the substance of these concerns and add three observations. First, we agree with Rabeyron that the very question of "where" psi originates, in participants versus in experimenters, is partially an ill-posed question once one has accepted the GQT framework. The ontic order from which acausal correlations emerge is not organized by the spatiotemporal coordinates that would license a clean source attribution. The experimenters, the participants, the

colors, and the apparatus form what Rabeyron (2026) terms a "communion of wills" within an entangled system. Asking where the autonomous impulse "is" within this system is analogous to asking where the wave is located in a wave–particle complementarity. Second, even if one were to grant the e-psi reading, the resulting phenomenon would still be a macroscopic non-classical correlation; what changes between scenarios (a) and (b) is the participant-versus-experimenter location of the autonomous impulse, not the acausal-versus-causal nature of the underlying coupling. Walach makes the same point in his commentary, and we take it to imply that the e-psi reading does not falsify the GQT interpretation but only relocates the autonomy within an already-entangled system. Third, we agree with Tierney that what our data cannot do (and what arguably no causality-testing experimental design can do) is determine which participant-versus-experimenter localization is the correct one. This is, in our view, exactly what should be expected if the GQT and the NT axiom are correct: a clean local-causal source attribution would itself constitute the kind of causal signal extraction that the NT axiom forbids.

The same logic applies to Rabeyron's (2026) Decision Augmentation Theory (DAT) reading of our results, in which the apparent erasure effect is reinterpreted as an unconscious precognitive selection of "good moments" by the experimenters. Within the GQT framework we have argued for, DAT and the macroscopic-complementarity reading are not mutually exclusive theories; they are two causal stories about an acausally connected system, neither of which can be definitively confirmed or refuted by a causality-testing design. The data are compatible with both, which is precisely what an acausal phenomenon should produce when forced into a causal interpretive frame.

On the Choice of Subjective and Objective Measurements

A brief comment on the choice of measurements. As we argued in the first part of this response, the central methodological precondition for the appearance of a macroscopic complementary relation is that the psychophysical correlation under study must contain an autonomous-subjective component. This is why we focused on motive-laden, preference-based subjective ratings, i. e., color likability in the original studies and food deliciousness in the food paper, rather than on purely descriptive subjective judgments such as brightness assessments. Likability and deliciousness are evaluative acts in which the perceiving subject's motives, preferences, and willing co-determine the response, and they thereby instantiate the induced-correlation type in the Pauli–Jung framework. Brightness assessments, by contrast, are descriptive perceptual judgments that primarily reflect a structural correlation between the objective stimulus property and its subjective registration; they should commute approximately normally and not yield an erasure effect.

The empirical contrast between these two classes of subjective measurements is informative. In Maier and Dechamps (2025), the brightness-variation results did not replicate beyond the initial two studies, whereas the likability effect appeared in the predicted direction across all four published studies and across the three additional studies reported here. The autonomy component thus appears to be the gating variable for the appearance of the effect at the first measurement level (individual perception), and the reduced-objective documentation strategy (the erasure-confound paradigm) is the gating variable for its survival at the second measurement level (scientific documentation across studies). On the objective side, the choice of measurements (the parametric color values and the food images' identifiers) was driven by the corresponding constraint: the objective values had to be stored or deleted from a results file without ever being inspected by experimenters or participants, so that the erasure manipulation could be executed cleanly under the conditions assumed by the GQT for non-locality to operate.

What the Data Show, and What They Cannot Show

We close with a summary. Our data cannot demonstrate autonomy directly, nor can they identify the source from which any autonomous impulse arises. What they can do, in conjunction with the parallel findings from the food domain reported in 2026, is constrain the alternative explanatory frameworks under which the observed pattern across the seven color studies and the four food studies could be accommodated. A purely physicalist account, which postulates the causal closure of a world describable through deterministic natural laws and direct-replication-validated regularities, predicts that macroscopic entanglement-based psychophysical interactions of the kind investigated here should not appear at all, or that any apparent appearance should rapidly dissipate under repeated documentation pressure. Neither prediction is supported by the data. To the extent that physicalism makes these specific predictions, the data declare physicalism incomplete in this empirical domain.

This is, we believe, the maximum that can be achieved with a deductive-causal method applied to a phenomenon whose ontic structure is, on the GQT account, inherently acausal. The deeper question asking what positive theoretical framework should replace physicalism given its incompleteness in this domain cannot, in our view, be answered with the same deductive approach. Tierney's (2026) suggestion that inductive methods may be required to characterize the alternative worldviews more positively is well-taken, and the conjecture he develops at the end of his commentary is one promising direction. The GQT in its current form is our preferred candidate framework. It is, however, only one of several candidate frameworks consistent with the data, and its full validation will require evidence of a kind that cannot, by the very logic

of the NT axiom, be delivered by any single deductive study. What our seven-study sequence does deliver is the more modest but nonetheless substantial result that a properly designed erasure-confound paradigm appears to be capable of documenting an acausal psychophysical correlation in a stable and reproducible way, without triggering the decline that has, until now, accompanied every other approach to this class of phenomena.

References

Aharonov, Y., & Vaidman, L. (2008). The two-state vector formalism: An updated review. In J.G. Muga, R. Sala Mayato, & I.L. Egusquiza (Eds.), *Time in quantum mechanics* (Vol. 1, pp. 399–447). Springer.

Atmanspacher, H., & Rickles, D. (2022). *Dual-aspect monism and the deep structure of meaning*. Routledge.

Atmanspacher, H., Römer, H., & Walach, H. (2002). Weak quantum theory: Complementarity and entanglement in physics and beyond. *Foundations of Physics, 32*(3), 379–406. https://doi.org/10.1023/A:1014809312397

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition, 2*(1), 16. https://doi.org/10.5334/joc.72

Edgington, E.S. (1972). An additive method for combining probability values from independent experiments. *Journal of Psychology, 80*(2), 351–363. https://doi.org/10.1080/00223980.1972.9924813

Fisher, R.A. (1932). *Statistical methods for research workers* (4th ed.). Oliver and Boyd.

Kane, R. (2002). Free will, determinism, and indeterminism. In H. Atmanspacher & R. Bishop (Eds.), *Between chance and choice: Interdisciplinary perspectives on determinism* (pp. 371–406). Imprint Academic.

Lucadou, W.v., Römer, H., & Walach, H. (2007). Synchronistic phenomena as entanglement correlations in generalized quantum theory. *Journal of Consciousness Studies, 14*(4), 50–74.

Maier, M.A., & Dechamps, M.C. (2025). Macroscopic complementary relation between subjective observations and objective measurements of colors. *Journal of Anomalistics, 25*(1), 15–60. https://doi.org/10.23793/zfa.2025.015

Maier, M.A., Dechamps, M.C., & Rabeyron, T. (2022). Quantum measurement as pragmatic information transfer: Observer effects on (s)objective reality formation. *Journal of Anomalous Experience and Cognition, 2*(1). https://doi.org/10.31156/jaex.23535

Maier, M.A., Vogel, A., & Dechamps, M.C. (2026). Macroscopic complementarity between subjective and objective food image assessments: A conceptual replication of the color-erasure effect. *Journal of Anomalistics, 26*(1), 37–74. https://doi.org/10.23793/zfa.2026.037

Maier, M.A., Vogel, A., Storch, A., & Dechamps, M.C. (2026). Non-classical correlation between subjective and objective color observations: Change of effect as a function of its empirical documentation. *Journal of Anomalistics, 26*(1), 14–36. https://doi.org/10.23793/zfa.2026.014

Rabeyron, T. (2026). Beyond the principle of elusiveness? Analysis of the experimental results of the erasure-confound paradigm. *Journal of Anomalistics, 26*(1), 75–96. https://doi.org/10.23793/zfa.2026.075

Römer, H. (2023). *Quanten, Komplementarität und Verschränkung in der Lebenswelt: Verallgemeinerte Quantentheorie*. Lit Verlag.

Schweizer, M.T. (2026). *Expectancy and reality: Perspectives on mind-matter-interactions* [Unpublished doctoral dissertation]. Ludwig-Maximilians-Universität München.

Squarra, N.F. (2025). *Beziehungen objektiver und subjektiver Variablen eines psychophysischen Systems: Eine Farb-Erasurestudie* [Unpublished Bachelor's thesis]. Ludwig-Maximilians-Universität München.

Tierney, I. (2026). Macroscopic complementary – questions and a conjecture. *Journal of Anomalistics, 26*(1), 97–112. https://doi.org/10.23793/zfa.2026.097

Walach, H. (2026). Chasing the wind: Empirical signatures of generalized entanglement. *Journal of Anomalistics, 26*(1), 113–119. https://doi.org/10.23793/zfa.2026.113

Walach, H., Kirmse, K.A., Sedlmeier, P., Vogt, H., Hinterberger, T., & Lucadou, W.v. (2022). Nailing jelly: The replication problem seems to be unsurmountable. Two failed replications of the matrix experiment. *Journal of Scientific Exploration, 35*(4), 788–828.

Walach, H., & Römer, H. (2000). Complementarity is a useful concept for consciousness studies. A reminder. *Neuroendocrinology Letters, 21*(3), 221–232.

Zusammenfassung

Akausale psychophysische Korrelationen in einer kausal geschlossenen Welt? Evidenz aus (nicht) konfundierten Löschungsparadigmen und die Grenzen physikalistischer Methodologie

Diese Erwiderung beantwortet die Kommentare von Thomas Rabeyron, Ian Tierney und Harald Walach zu zwei Artikeln diesem Heft über makroskopische komplementäre Beziehungen zwischen subjektiven und objektiven Aspekten der Realität im Rahmen der Verallgemeinerten Quantentheorie (VQT). Die zentrale gemeinsame Frage der Kommentatoren lautet, ob das Erasure-Confound-Paradigma den vom Non-Transmission-(NT-)Axiom postulierten Effektrückgang, der frühere experimentelle Paradigmen in diesem Forschungsfeld betroffen hat, dauerhaft vermeiden kann. Wir argumentieren, dass das NT-Axiom eine epistemische

Beschränkung ausdrückt, die immer dann auftritt, wenn eine kausal-deterministische Dokumentation auf Phänomene mit einer autonom-subjektiven Komponente angewendet wird. Unter dieser Lesart erreicht das Erasure-Confound-Paradigma auf der Einzelstudienebene nur eine reduziert-objektive Dokumentation des Effekts, weil der Confound eine eindeutige kausale Interpretation verhindert, während der Effekt studienübergreifend indirekt über den Ausschluss der alternativen Erklärung (des „Color-Selection-Bias"-Confounds) etabliert wird. Dieses Design entkoppelt die Evidenzakkumulation vom Zeitpfeil und soll dadurch den Effektrückgang umgehen. Anschließend prüfen wir die Rückgangsvorhersage empirisch, indem wir drei bislang unveröffentlichte Farb-Erasure-with-Confound-Studien (Studien 5–7) berichten, die zusammen mit den vier Studien von Maier und Dechamps (2025) den vollständigen Datensatz von sieben Studien in diesem Paradigma bilden, ohne Schubladenproblem. Der Erasure-Effekt zeigte sich in jeder Studie in der vorhergesagten Richtung (Vorzeichentest $p = .008$), und kombinierte stratifizierte Permutationsanalysen über drei unabhängige 36-Farben-Sets schlossen den Color-Selection-Bias-Confound als gemeinsame Erklärung aus (Fisher $p = .010$; Edgington $p = .018$). Eine inverse-varianz-gewichtete Meta-Regression über die sieben Effektstärken ergab einen nicht signifikanten negativen Gradienten ($b = -0{,}018$, 95%-KI [−0,043, 0,006], einseitig $p = .07$), und die siebte Studie zeigte einen weiterhin deutlich vorhandenen Effekt ($BF_{10} = 3.679{,}73$, $d = .30$). Die Daten sind mit einer stabilen, reduziert-objektiv dokumentierten akausalen psychophysischen Korrelation vereinbar und nicht mit dem vom NT-Axiom unter direkter Replikation vorhergesagten Effektrückgang. Weitere von einzelnen Kommentatoren aufgeworfene Punkte, etwa zur Experimenter-Psi-Interpretation und zur Wahl der subjektiven und objektiven Messungen, werden in kürzerer Form behandelt.

Schlüsselbegriffe: Verallgemeinerte Quantentheorie, Non-Transmission-Axiom, makroskopische Komplementarität, Erasure-Confound-Paradigma, psychophysische Korrelationen, Effektrückgang, akausale Korrelationen

Journal of Anomalistics
Zeitschrift für Anomalistik

www.anomalistik.de www.igpp.de

Probing Top Performers in a Forced-Choice Clairvoyant Task

Helané Wahbeh [a,*], Michael Kriegsman [a], Beth Glick [b], Arnaud Delorme [a], Dean Radin [a]

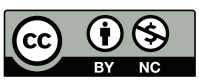
Abstract – This preregistered study analyzed more than 25 million trials from a web-based forced-choice remote viewing task to examine patterns of clairvoyant performance across all participants and among a subset of top performers. At the aggregate level, runs ending at the four planned lengths (5, 10, 25, or 100 trials) conformed to chance expectations. In contrast, optionally stopped runs displayed systematic fluctuations: short runs (1–3 trials) were above chance before declining, runs of 11–19 fell below chance, and runs beginning at 20 showed recurring above-chance spikes at every fifth run length (e.g., 30, 35, 40, 45, 50), which diminished beyond 80 trials. A Monte Carlo simulation matched to the empirical stopping distribution clarified the extent to which these patterns could be reproduced by optional-stopping behavior alone, with much of the run-length pattern, including the 11–19 trough and round-number variability, falling within the simulated null envelope. Exploratory analyses of top performers, defined post-hoc as the 1,235 users (2.64%) who exceeded chance uncorrected, after no users survived the preregistered FDR-corrected criterion, examined belief in psi, prior precognitive experience, meditation, total trials, and optional stopping as predictors. Optional stopping was the predictor most consistently associated with hits, both on first trials and across all trials, where it also interacted with belief, prior precognitive experience, and meditation jointly with cumulative task experience. Effect sizes were small (Δp and Cohen's d near zero for most predictors), and findings are interpreted as exploratory. The results suggest that group-level outcomes primarily reflect optional-stopping and related behavioral dynamics, whereas top-performer analyses surface more nuanced,

(a) Research Department, Institute of Noetic Sciences, Novato, California

(b) California Institute for Human Science, Encinitas, California

* Corresponding author: Helané Wahbeh

hwahbeh@noetic.org



but small, context-dependent associations between belief, experience, and behavior. These findings highlight the methodological challenges of large-scale, open online testing and the value of preregistered, participant-level approaches, combined with benchmark simulations, for distinguishing behavioral artifacts from potential psi signals.

Keywords: clairvoyance, forced-choice task, optional stopping, top performers, individual differences, preregistration

Introduction

Forced-choice tasks are a central experimental paradigm in parapsychology, requiring participants to choose among fixed alternatives to determine whether correct guesses exceed chance expectations. These designs afford precise control and have produced statistically significant, though small, effects in meta-analyses (Baptista et al., 2015; Cardeña, 2018; Storm et al., 2010; Storm & Tressoldi, 2020; Tressoldi & Storm, 2024a). However, their reliance on aggregating trials across participants has attracted methodological scrutiny.

Traditionally, parapsychological researchers have assumed that all trials are independent, enabling the use of straightforward statistical models. Under the null hypothesis, this assumption holds, and pooled trials increase statistical power and the apparent robustness of findings (Palmer, 1985). However, critics have noted that pooling can obscure important individual differences and inflate significance in the presence of statistical dependencies (Hyman, 1995; Utts, 1996). Because forced-choice tasks are among the most frequently cited evidence for psi, unexamined dependencies in trial data have implications for both interpretation and the credibility of the field[1].

Empirical and theoretical work suggests that the independence assumption may not always be justified. Participant performance can be influenced by factors such as learning, fatigue, feedback, and expectation bias, introducing potential dependencies between sequential trials (Dalkvist et al., 2014; Kennedy, 2013, 2014; Varvoglis et al., 2019). Failure to account for statistical dependence may lead to spurious results, exaggerated effect sizes, and invalid conclusions (Kennedy, 2016).

Moreover, there is substantial individual variability in psi performance. Selected participants, or those with traits, beliefs, or backgrounds thought to favor psi, tend to achieve higher

1 While free-response tests typically have stronger effect sizes, forced-choice tests have been commonly used in parapsychological research (see Storm et al., 2010, 2012; Storm & Tressoldi, 2020 for meta-analyses of forced choice and free-response studies).

hit rates than unselected samples (Cardeña, 2018; Storm et al., 2010; Tressoldi & Storm, 2024a; Utts, 1996). For example, Tressoldi and Storm (2024a) found that individuals selected for characteristics or experiences thought to be linked to psi performance, such as prior experience with extrasensory perception (ESP) experiments, belief in psi, psi training, or long-term meditation practice, showed effect sizes up to three times higher than unselected participants. These differences underscore the need for participant-level analyses that explicitly model within-participant variability and incorporate psychological and contextual moderators. To confront these challenges, parapsychological researchers have begun advocating for participant-level analyses, calculating individual hit rates, and explicitly modeling within-subject variation (Kennedy, 2013, 2016). Such methods offer a more nuanced and accurate understanding of psi effects, mitigating the limitations of aggregate analyses.

The Current Study

Beginning in 2000, Radin et al. (2019) launched the web-based GotPsi experiment suite, which by 2025 had collected hundreds of millions of trials from over 300,000 participants worldwide and produced significant results for some tests. In 2005, a new task called "Quick Remote Viewing" (QRV) was introduced, a misnomer, as it is best described as a simple forced-choice clairvoyance task. Building on previous research, the present study conducts preregistered analyses addressing two questions: (1) how hit rate varies as a function of run length across all participants, and (2) among participants who perform significantly above chance, which psychological and behavioral factors best predict performance, both for first-trial success and across all trials.

By combining a large archival database (over 25 million trials) with participant-level analyses of top performers, this study aims to clarify the extent of statistical dependence in forced-choice psi tasks and identify individual differences linked to exceptional performance.

Specifically, we address the following research questions that are approached with exploratory analyses:

Research Question 1. How does hit rate vary as a function of run length across all participants?

Research Question 2. Among participants who perform significantly above chance on the QRV task:

- 2a. Which, if any, of the following predictors are associated with first-trial success: belief in psi, past precognitive experience, meditation/meditative movement, total number of trials completed, and optional stopping ratio (including all main effects and interactions)?
- 2b. Which of these factors, if any, best explain success across all trials completed?

By simultaneously examining group-level trends and individual variability, this study aims to provide a more detailed understanding of the predictors and mechanisms that may support above-chance performance in forced-choice psi tasks. These analyses were preregistered with the Koestler Parapsychology Unit (KPU Registry 1094). Our preregistration specified these analyses, including the complete set of predictors and interactions. Here we present them as broader research questions to aid readability, while adhering to the pre-registered plan.

Methods

Study Procedures

Participants registered on the GotPsi website, www.gotpsi.org. During registration, they entered information about themselves, including handedness, beliefs, and experiences of psi, creativity, and remote viewing training and experience. They could then choose to engage in one of eight tasks. This study specifically examines the Quick Remote Viewing (QRV) task, which was initiated within the task suite on April 8, 2005. All human study activities were approved and overseen by the Institutional Review Board at the Institute of Noetic Sciences (IORG#0003743).

Task Description

The QRV task is a forced-choice photo guessing task where the participant attempts to identify a target image selected by the web server from five possible choices. For this task, a blank frame is displayed in the center of the screen, and five photos, drawn from a pool of 130 images, are displayed below it. The images were selected from the Corel Professional Photo image database and depict various scenes, including individuals, nature, and urban scenes. When the participant begins a trial, the web server selects the target image using a Perl-based linear congruential pseudorandom number generator function. This target exists in server-side state before the participant views the five response options and makes their selection. The participant then selects which of the five images they believe matches the hidden target, after which the target is revealed in the blank frame. Because the target is determined at the start of each trial rather than after the participant's choice, the QRV task is best characterized as a real-time clairvoyance paradigm (the target exists at the time of response and the participant attempts to identify it), rather than a precognition paradigm (in which the target would not yet be determined at the time of response). Although the study was preregistered under the label "precognition,"

subsequent clarification of the task code established that clairvoyance is the more accurate characterization. When the target picture is displayed, the participant sees if their choice was correct or not, and the following feedback is displayed: "That was a hit" or "That was a miss." The participant then presses a button to continue to the subsequent trial. The participant can select the number of trials they would like to perform, 5, 10, 25, or 100, which we define as the run length (run_length) for this study. Ten trials was the default option; users who did not actively select a different run length would begin a 10-trial run. Despite selecting the number of trials the participant intends to provide for each run, they could discontinue at any time. Because the task offered only four planned run lengths (5, 10, 25, and 100), any run ending at other lengths greater than 25 necessarily reflects within-run stopping from a planned 100-trial attempt rather than a chosen planned length. However, the dataset records only the number of trials actually completed per run, not the participant's chosen planned length. Runs ending at 5, 10, or 25 trials therefore cannot be distinguished between completed planned runs and optionally stopped attempts from a longer planned length. Only runs ending at 100 trials can be unambiguously identified as completed planned runs. If the participant completes the run, the participant is shown the percentage of "hits" achieved. A *z*-score is assigned to each run. The *z*-score is calculated as follows,

$$Z = \frac{(h - np)}{\sqrt{np(1-p)}}$$

where n is the number of trials in the run, p is the probability of hitting a target (0.2), h is the number of hits or correct responses, np is the mean hit rate, and $\sqrt{np(1-p)}$ is the standard deviation of a binomial distribution. For the present dataset, participants were able to take and repeat the task at their leisure. Thus, data can be viewed at multiple levels: as one omnibus hit rate, or as nested hit rates: within participant, within year/date, or within observed run length.

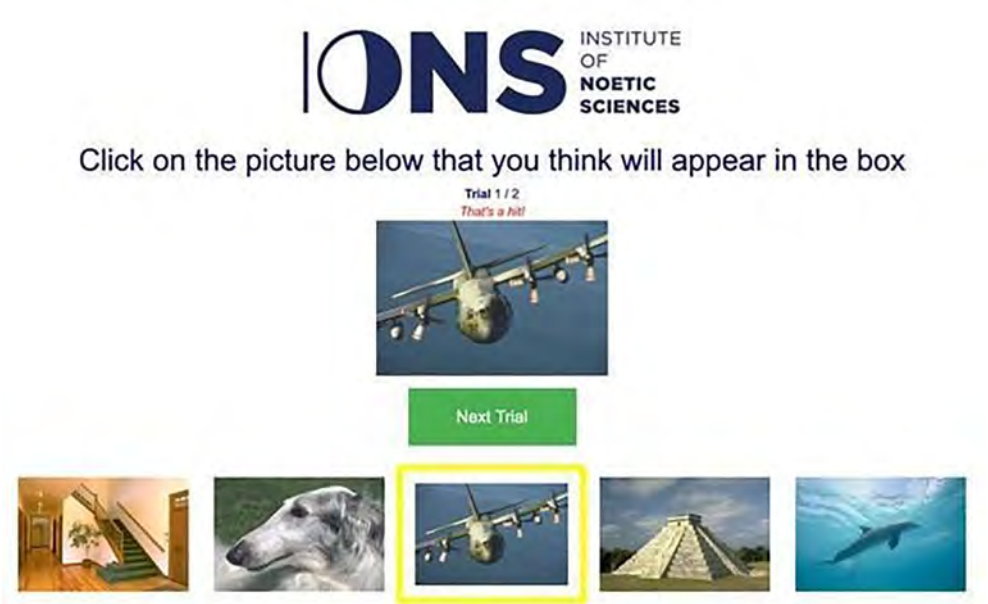
Subjective Measures

Basic demographics such as age, gender, race, and socioeconomic status were not collected for any participants. Handedness, creativity, and remote viewing experience were completed and not included in these analyses (see Supplemental Data for these methods and results).[2] Participant Beliefs and Experiences were part of the current analyses, and the measure is described below.

2 The Supplemental Data are available at: https://www.anomalistik.de/images/pdf/zfa/supp_mat/JAnom26-1_QRV-Supplemental-Materials.pdf

Figure 1

QRV Participant View



Note. This is the participant's view of the task after they have made their selection and the correct image has been revealed (highlighted in yellow). In this case, the trial was a hit.

Beliefs and Experiences (Table 3) – Participants were asked, "Please answer the following questions using a 5-point scale. For example, your answer to the question, "The degree to which you practice meditation," can range from "none" (the leftmost radio button) to "extensive" (the rightmost button)." These items could be considered ordinal. However, we considered them continuous because they were displayed to the user as equal intervals (i.e., the steps or gaps between each point on the scale are equal, and the perceived difference in agreement or frequency between each level is consistent across the scale). Items 1, 2, 3, 4, 12, and 13 were included in the regression models as potential predictors. Item 1 is about "*Belief*." Items 2, 12, and 13 are averaged to create a variable called "*Precog*." Items 3 and 4 are averaged to create a variable called "*Meditation*."

Data Cleaning

The present data set is extraordinarily vast because participants were able to repeat the task *ad infinitum*. Raw data were collected from April 8, 2005, to the present (October 2025). However, the data analyzed for this study included trials contributed only through December 29, 2018, because at that point, the underlying code, database, and host server were updated using more modern web-based programming methods. The QRV task was modified with the 2018 system update, and although data collection has continued with the updated platform through the present, the task changes made merging the pre- and post-update datasets inappropriate for the present analyses. The post-2018 data are retained for future analyses and could serve as an independent dataset for confirming the exploratory findings reported here. The raw data were imported into R (R Development Core Team, 2023) and merged into a single file for cleaning and analysis. Rows, each corresponding to one trial, were reordered by Username and Date to organize consecutive trials within their corresponding runs. A cleaning algorithm was then applied to flag, triage, and remove malformed data records and potentially nefarious attempts to game the system, such as utilizing multiple windows or bots. Specifically, the following types of errors were addressed: NAs (no discernable value due to unknown cause), disk access glitches causing duplicate consecutive trial numbers, skipped trial numbers, trials submitted with identical timestamps, or trial or hit count carryover into subsequent runs, and trials submitted "too fast" (1 second or less elapsed), suggesting that those trials were performed by bots attempting to hack into the web server. Among the raw data (i.e., before other cleaning steps), 0.51% of trials had a recorded duration of 0 seconds, and 16.36% had a duration of 1 second or less; inter-trial intervals are recorded as integer timestamps and therefore lack finer-grained resolution. Our rationale for this 1 second threshold is that participants must (1) view five target images, (2) click one, (3) see the feedback/result, and (4) proceed to the next trial, which we posit cannot be completed legitimately in under 1 second.

NAs were found only in the trial number variable, and only between runs. More specifically, NAs seemed to appear between runs that were submitted on different days, suggesting an encoding issue when the browser window was left open overnight. Because NAs occurred only on trials between runs, the implemented cleaning strategy was to remove these rows. Duplicated consecutive trial numbers were first segmented into two types. Consecutive first trials could occur when participants answered one trial, then quit and restarted a new run. This amounts to optional stopping, not an explicit error. However, consecutive trial numbers larger than 1 did indicate some kind of error. Trial skips were defined as instances where one or more trial numbers were missing in the consecutive count of trial numbers within a run. Carryover hits corresponded to the impossible situation where the hit count exceeded the trial number.

Carryover hits also included instances where the first trial was a miss, and yet the hit count was 1. These checks were applied to every trial in the dataset. In addition to the two patterns described above, runs were also flagged if a trial with trial number = 1 had hit = 1 but hit count > 1 (i.e., the first trial was a hit but the count had already carried over from a prior run). All retained trials satisfy hit count ≤ trial number.

These errors could have various causes, including users attempting to game the system (e.g., running the task with multiple windows open simultaneously, programming bots, or rapidly submitting responses), or errors in the back-end encoding of the data collection software. Because we could not discern the actual cause of these errors, we opted to remove runs that contained flagged trials with these error types (i.e., all trials for the suspect run).

Statistical Plan

For trial-level analyses, a random subset of 25% of users across all years was withheld for unplanned future analyses. The purpose of this withholding was to preserve a portion of the dataset for future work, rather than to support a specific planned analysis. In particular, because the top-performer analyses reported here are exploratory, the held-out subset could serve as an independent sample for confirming, or failing to confirm, the effects observed in the present analyses. Statistical analyses included three sections: (a) descriptives, (b) run-level analyses exploring Research Question 1, and (c) trial-level analyses exploring Research Question 2.

Descriptive statistics (means, *SDs*) were computed and two-tailed Welch's *t*-test was used to compare the 15 belief and experience variables across Top Performers versus Not Top Performers (see Table 3). Means and standard deviations for continuous variables, as well as counts and percentages for categorical variables, were calculated and presented for all included participants, including the random subset and the subset of top performers (see below). Participant characteristics were compared between top performers and non–top performers using Welch's two-sample *t*-tests, with effect sizes estimated by Cohen's *d* and accompanied by 95% confidence intervals.

Research Question 1:

How does the hit rate vary as a function of run length in all participants? (exploratory hypotheses with no *a priori* prediction)

The effect of optional stopping and patterns of performance by run length were assessed using a ribbon plot (see Figure 2), which shows the distribution of hit rates on the y-axis and the

number of trials in each given run on the x-axis. *Post-hoc* binomial tests were also conducted at each of the 100 run lengths to test if the average hit rate differed from chance. In addition to hit rates, we plotted the log of the number of runs submitted at each length. This was not preregistered but was included to aid interpretation, as run counts varied greatly across lengths. The log transform allowed both common and rare run lengths to be visualized on the same scale, providing context for evaluating which deviations from chance occurred at highly frequent versus rare run lengths. Descriptive binomial tests of hit rate were conducted for six nested trial subsets (1st trial, 1st 5, 1st 10, 1st 25, 1st 100, and all trials), separately for the full 75% retention sample and for the Top Performer subset. This nested-subset framing was consistent with the preregistered Stats Plan, which specified that hit rates could be examined at multiple levels. Finally, following the reviewer's suggestion, we conducted a Monte Carlo simulation to evaluate whether the observed run-length patterns could be explained by optional stopping alone. Null data were generated by drawing hits from a binomial distribution with $p = 0.20$, using run lengths sampled from the empirical run-length distribution of the 75% retention sample. The simulation was repeated 1000 times to construct a 95% null envelope against which the observed mean hit-rate curve (Figure 3) could be compared. Deviations of the observed curve beyond the envelope indicate patterns not reducible to optional-stopping behavior under the null.

Research Question 2:

Identifying Top Performers

In the preregistration, we intended to define top performers as those users who performed significantly above chance after FDR correction (Benjamini & Hochberg, 1995). No users survived FDR correction, so we post-hoc modified the definition of top performers to correspond to the 1235 users (2.64% of the total 46,722 users) who performed above chance, uncorrected. The full distribution of user-level Z scores across the sample is shown in Supplemental Figure 1, with corresponding numerical comparisons against the standard normal distribution in Supplemental Table 9; the observed upper tail of the distribution exceeded normal-theory expectations at every threshold examined (e.g., 9.06% of users exceeded $z = 1.645$ versus 5.00% expected).

Research Question 2a:

Which, if any, of the five predictors, belief in psi, past precognitive experience, meditation/meditative movement, total number of trials completed, and optional stopping ratio (including all main effects and interactions), are associated with top performers' success on their first

trial (i.e., first trial ever completed by each user)? The five predictors were chosen based on what the archival registration data actually measured and on prior literature. Belief in psi, prior precognitive experience, and meditation were included as trait-level predictors because each has been linked to psi performance in prior work (Cardeña, 2018; Storm et al., 2010). Total trials and optional stopping ratio were included as behavioral predictors of task engagement and stopping dynamics, with optional stopping, in particular, being repeatedly implicated as a driver of apparent above-chance effects. Remaining registration items (handedness, creativity, and remote-viewing training) did not have the same theoretical grounding and are reported descriptively in the Supplement. The first trial of each user's first run was used because it represents a unique instance where every user had no prior experience with the task, providing a homogenized dataset for cross-user comparison. This choice also minimizes contamination from optional stopping (since no stopping decisions have yet been made), yields one independent observation per user, and is consistent with prior reports of front-loaded or early-trial psi effects (Kennedy, 2003; Mossbridge & Radin, 2018). Hit outcomes were modeled with logistic regression, with a logit offset for chance ($p = 0.2$). Two models were estimated: a main-effects-only model (Supplemental Table 5) and the preregistered interaction model (Supplemental Table 6), which includes pairwise and three-way interactions among the three Psi-type predictors and Total Trials × Optional Stopping (see Supplement for full list of terms).

Research Question 2b:

Which of these factors, if any, best explain success across all of their trials? (Exploratory hypotheses with no *a priori* prediction.) Four regression-based models were assessed, all with the same five continuous predictors: (a) belief in psi (Belief), (b) past precognitive experience (Precog), (c) meditation and meditative movement (Meditation), (d) total number of trials completed (TotalTrials), and (e) optional stopping ratio (OptStop): the number of trials from runs optionally stopped divided by total number of trials. For each outcome (first trial; all trials), two models were fit: a main-effects-only model and the preregistered interaction model that adds pairwise and three-way interactions among the predictors (see Supplement for full list of terms). The first pair of models uses logistic regression to predict each top performer's first trial (hit or miss). The second pair predicts all trials from all top performers. In addition to reporting statistical significance, we calculated delta probabilities (Δp), which represent the average marginal change in predicted probability of a hit for a one-unit increase in each predictor. Because our dataset is very large, p-values alone could indicate significance for substantively trivial effects; Δp provides a more interpretable measure of effect size on the probability scale.

To control Type I error, FDR correction (Benjamini & Hochberg, 1995) was applied to the non-intercept predictors within each logistic regression model. This conservative approach exceeds standard practice but ensures that reported effects remain reliable even under a broad family-wise correction. All models were estimated as fixed-effect logistic regressions, as preregistered.

Results

Data Collected and Cleaning Procedures

There were 26,708,074 raw data trials collected from 64,775 participants (i. e., each unique username was treated as an independent participant, although a single person may have created multiple usernames). There were 25,457,187 trials after data cleaning. Please see Supplemental Data Table 2 for detailed trials per year and Supplemental Data Table 3 for the number of trials and users by year after cleaning procedures were implemented over all participants. The number and percent of trials removed during cleaning were as follows (see Methods): NAs = 8,736 (0.033%); duplicated consecutive trial numbers = 125,233 (0.469%); skipped trial numbers = 9,397 (0.035%); carryover hits = 16,269 (0.061%); and too fast = 211,271 (0.791%).

All top-performer analyses used N = 1,235 users (defined post-hoc as those exceeding chance uncorrected; see Methods) contributing 414,951 runs and 4,432,173 trials.

Table 1

Number of Trials, Runs, and Users Included in Analyses

	Trials	Runs	Users
All participants			
Raw data	26,708,074	not computed	64,775
Cleaned data	25,457,187	2,479,188	62,296
25% Users withheld	18,122,863	1,655,523	46,722
Top Performers	4,432,173	414,951	1,235
Not Top Performers	13,690,690	1,240,572	45,487

Table 2.

Descriptive hit rates by trial subset for the full 75% retention sample and Top Performers

Trial subset	Users (*N*)	Trials	Hits	Hit rate	Binomial *p*	Weighted mean	95% CI
Panel A. Full 75% retention sample							
1st trial	46,722	46,722	9,658	0.207	< .001	0.207	[0.203, 0.210]
1st 5 trials	36,615	183,075	39,179	0.214	< .001	0.214	[0.212, 0.216]
1st 10 trials	22,004	220,040	47,612	0.216	< .001	0.216	[0.215, 0.218]
1st 25 trials	11,053	276,325	58,307	0.211	< .001	0.211	[0.210, 0.213]
1st 100 trials	3,069	306,900	63,082	0.206	< .001	0.206	[0.204, 0.207]
All trials	46,722	18,122,863	3,633,161	0.201	< .001	0.201	[0.200, 0.201]
Panel B. Top Performers (exploratory, post-hoc; *N* = 1,235)							
1st trial	1,235	1,235	615	0.498	< .001	0.498	[0.470, 0.526]
1st 5 trials	1,128	5,640	2,453	0.435	< .001	0.435	[0.422, 0.448]
1st 10 trials	870	8,700	3,279	0.377	< .001	0.377	[0.367, 0.387]
1st 25 trials	497	12,425	3,937	0.317	< .001	0.317	[0.309, 0.325]
1st 100 trials	151	15,100	3,782	0.251	< .001	0.251	[0.242, 0.259]
All trials	1,235	4,432,173	900,881	0.203	< .001	0.203	[0.202, 0.204]

Note. Hit rate expected under chance = 0.200. For each subset, the number of users is the count of participants who contributed at least the indicated number of trials. Binomial *p*-values test whether the observed hit rate differs from 0.200. The weighted mean accounts for each user's contribution by total trials; over users, averages converge to the same omnibus rate across all trials. Panel B presents the same analyses restricted to the post-hoc Top Performer subset (users exceeding chance uncorrected; see Methods).

Descriptive hit rates by trial subset are presented in Table 2 for both the full 75% retention sample and the Top Performer subset. In the full sample, hit rates were slightly above chance across all subsets (ranging from 0.201 to 0.216) and highly significant given the very large sample sizes, though effect magnitudes were small. Among Top Performers, hit rates were substantially elevated and declined monotonically as more trials were included: from 0.498 on the first trial alone, to 0.435 across the first 5 trials, 0.377 across the first 10, 0.317 across the first 25, 0.251 across the first 100, and 0.203 across all trials. This pattern is consistent with a front-loaded effect but also with selective retention, because Top Performers were defined by above-chance performance overall, the pattern of early elevation may partly reflect this selection rather than a genuine front-loaded effect.

Participants Characteristics

Means, standard deviations, and group comparisons with two-tailed Welch's two-sample *t*-tests are presented in Table 3. To provide descriptive transparency, the table summarizes participants' self-reported beliefs and experiences separately for Top and Not Top performers. Although these comparisons were exploratory and not preregistered, they offer context for understanding whether the Top Performer subset differed meaningfully on background variables. Of the 15 self-report measures, two showed statistically significant differences after FDR correction. Top performers reported slightly more frequent meditation practice than non–top performers. Top performers also rated their view of time slightly closer to "raging waterfall" than "placid pool" (see Table 3). Although these differences were statistically reliable, the effect sizes were very small ($|d| \approx 0.07$, with confidence intervals extending nearly to zero), and are unlikely to reflect substantively meaningful psychological distinctions.

Research Question 1: Hit Rate Relationship With Run Length in All Participants

The ribbon plot illustrating the relationship between hit rate and run length is displayed as Figure 2. This analysis examines aggregate hit rates as a function of run length. Because the task offered only four planned run lengths 5, 10, 25, or 100 trials, runs ending at other lengths above 25 necessarily reflect within-run stopping from planned 100-trial attempts; runs ending at 5, 10, or 25 may represent either completed planned runs or optionally stopped attempts from longer planned lengths (see Task Description). Thus, any runs consisting of a different number of trials corresponded to a run in which the participant elected to stop the run before completion (or there was a glitch in the data encoding that stopped the run). Results for runs ending at the four planned lengths (5, 10, 25, and 100 trials) conformed to chance: none differed significantly from 0.2 in the run-level analysis underlying Figure 2 (Supplemental Table 4a). The trial-level binomial tests in Supplemental Table 4b flagged lengths 10 and 100 as significant, but the corresponding deviations were negligible (hit rates of 0.1994 and 0.2007; Cohen's $h \approx 0$), reflecting the very large number of trials at these common stopping points rather than a meaningful departure from chance.

Table 3

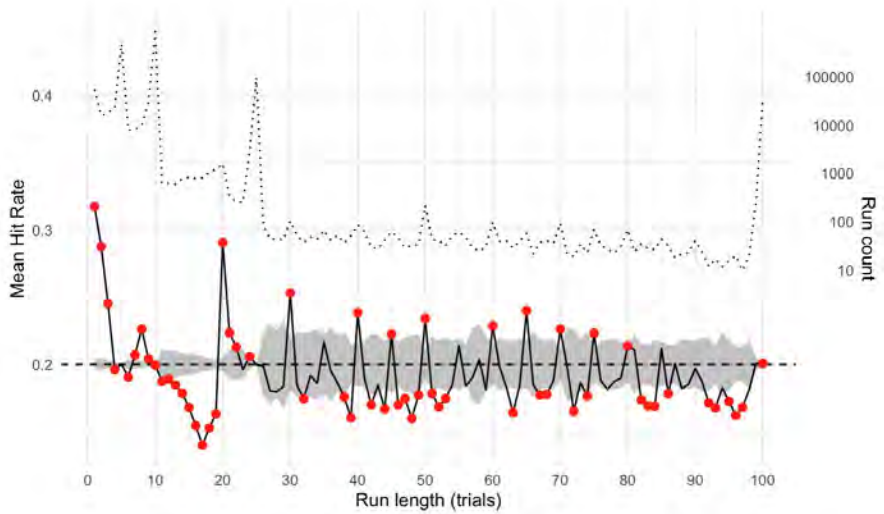
Participant Beliefs and Experiences

Degree to which you…	NOT Top Performers *N* = 45,487		Top Performers *N* = 1,235		Welch's Two-Sample *t*-test		Cohen's *d* [95% CI]
	Mean	*SD*	Mean	*SD*	*df*, *t*	*p*	
1. believe in "psychic" phenomena (none, absolute)	4.23	0.99	4.24	0.93	1200, -0.44	.66	-0.01 [-0.07, 0.05]
2. have had precognitive experiences (none, extensive)	3.45	1.16	3.48	1.13	1194, -0.75	.45	-0.02 [-0.08, 0.04]
3. practice meditation (none, extensive)	2.57	1.28	2.66	1.30	1188, -2.24	.03*	-0.07 [-0.31, -0.01]
4. practice martial arts or yoga (none, extensive)	2.03	1.25	2.09	1.29	1185, -1.65	.10	-0.05 [-0.11, 0.01]
5. consider yourself creative (none, extremely)	3.89	1.03	3.90	1.01	1192, -0.35	.72	-0.01 [-0.07, 0.05]
6. are lucky (none, extremely)	3.21	1.11	3.23	1.07	1195, -0.91	.36	-0.03 [-0.09, 0.03]
7. trust in your intuition (never, always)	3.95	0.90	3.98	0.86	1196, -1.30	.20	-0.04 [-0.10, 0.02]
8. trust in a religious faith (none, absolute)	2.86	1.48	2.89	1.46	1193, -0.81	.42	-0.02 [-0.08, 0.03]
9. have a sense of the spiritual (none, absolute)	3.99	1.12	4.00	1.11	1191, -0.35	.72	-0.01 [-0.07, 0.05]
10. are enthusiastic about sports (none, extremely)	2.59	1.37	2.62	1.37	1192, -0.53	.60	-0.02 [-0.08, 0.04]
11. work as a scientist (none, extensive)	1.83	1.21	1.80	1.18	1193, .04	.30	0.03 [-0.03, 0.09]
12. are trained in remote viewing (none, extensive)	1.32	0.80	1.32	0.81	1188, -0.07	.95	0.00 [-0.06, 0.06]
13. actively remote viewing (none, extensive)	1.43	0.92	1.48	0.96	1183, -1.42	.16	-0.04 [-0.10, 0.01]
14. have participated in psi experiments (none, extensive)	1.54	1.02	1.55	1.05	1180, -0.30	.77	-0.01 [-0.07, 0.05]
15. view time metaphorically as a (placid pool, raging waterfall)	3.01	1.20	3.10	1.21	1187, -2.38	.02*	-0.07 [-0.13, -0.01]

Note. Groups compared via Welch's two-sample *t*-test, equal variances not assumed, applying Satterthwaite approximation for degrees of freedom. Reported *p*-values are unadjusted. Asterisks indicate predictors that remained significant after false discovery rate (FDR) correction for multiple comparisons ($p < .05$). The *N* for Top Performers differs from the *N* in Table 1 because not all participants opted to complete these items.

Figure 2

Ribbon Plot of Hit Rate as a Function of Run Length



Note. In this ribbon plot, each value along the X axis corresponds to all runs submitted for that given run length. The solid line shows the mean hit rate on the left-sided y-axis, and circles indicate run lengths where the mean hit rate was significantly different from 0.2 after FDR correction for these 100 tests. The gray ribbon corresponds to the null envelope, or mean hit rate of 0.2 +/- 1.96 SE, FDR corrected. The dotted line shows the log of the number of runs submitted for each given run length, indicated by the right-sided y-axis.

The ribbon plot in Figure 2, with significant values listed in Supplemental Table 4, reveals five patterns in the average hit rate (represented by the black line), which are outlined in the following five paragraphs. In brief, these five patterns include that the mean hit rate (1) was nonsignificant for complete runs, (2) decreased continually over run lengths 1–4, (3) 11–19, and (4) 65–97, and (5) spiked upward on every fifth run length beginning at 20. In addition to hit rates, Figure 2 includes the log of the number of runs submitted at each length (dotted line), providing context for how often each run length occurred. This allows deviations from chance to be interpreted in light of the frequency with which participants stopped at those run lengths. As expected, there were much greater run counts at 5, 10, 25, and 100.

Pattern 1: At the four planned run lengths 5, 10, 25, or 100 trials, the mean hit rate was not significantly different from 0.2, though it should be noted that runs ending at 5, 10, or 25 are a mixture of completed planned runs and runs optionally stopped at those lengths from longer

planned attempts. Deviations from chance emerged primarily at run lengths outside the planned options, which necessarily reflect optional stopping. The log of run counts revealed four pronounced spikes at these planned run lengths, confirming that they were by far the most common stopping points. By contrast, incomplete runs showed a much more variable distribution of frequencies, with some lengths rarely attempted and others clustering around round numbers. These patterns provide important context for interpreting the fluctuations in hit rate described below.

Pattern 2: The second pattern was a decreasing hit rate for run lengths 1–4. Overall, runs consisting of only one, two, or three trials showed hit rates significantly above chance, with a descending magnitude. This pattern continued for runs with four trials, where the mean hit rate is significantly less than 0.2. Runs with six and eight trials were significant, but no clear pattern was observed for run lengths between six and ten trials.

Pattern 3: Run lengths from 11 to 19 were significantly lower than chance, with a general pattern of decreasing hit rate. This pattern reversed dramatically at runs with 20 trials, for which the average hit rate was significantly above chance.

Pattern 4: Run lengths from 65 to 97 also revealed another pattern of decreasing hit rates. The overall pattern observed for these run lengths above 65 can be seen as the additive effects of patterns 1, 4, and 5 (discussed below).

Pattern 5: A recurring upward spike in hit rate was observed beginning at runs of 20 trials and reappearing at every fifth run length thereafter (e.g., 30, 35, 40, 45, 50, etc.). The log of run counts showed the same periodic pattern, indicating that these upward deviations in hit rate coincided with points where participants disproportionately chose to stop. In other words, the peaks in performance were not randomly distributed but systematically aligned with round-numbered stopping points, suggesting that stopping behavior was closely linked to the observed fluctuations in hit rate.

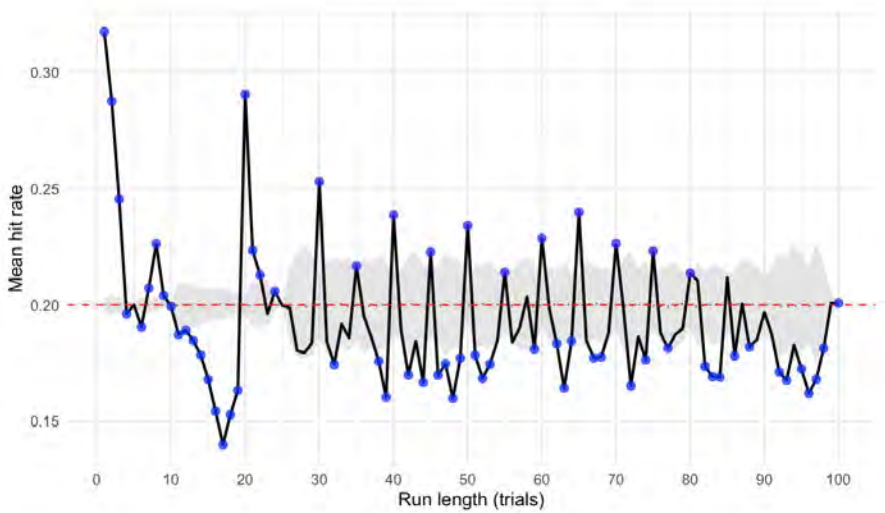
Finally, there were overlapping additive effects of the above patterns. Of note, the upward spiking of the Pattern 5 was not seen for runs of length 25 or 100 (see the Pattern 1), but was observed for the remainder of trials (i.e., 20, 30, 35, 40, 45, 50, etc.). For run lengths above 65, fluctuations in hit rate followed both the gentle decrease in hit rate of Pattern 4 and the spiking of Pattern 5. This blending of patterns was most interesting for run lengths between 92 and 97, where the positive spiking was observed for run lengths of 95 (and surprisingly, more so for 94). However, the magnitude of the decreasing hit rate overpowered the spike, so that even with the positive spike in hit rate at 95, the hit rate there was in fact significantly below chance. In other words, all run lengths between 91 and 98 were significantly below chance (except 94), such that the positive spiking at 95 was masked by the decreasing trend from 65 to 97.

To summarize, when aggregated across all participants, completed runs of 5, 10, 25, or 100 trials showed no significant deviations from chance expectation ($M \approx 0.20$ hit rate). In optionally stopped runs, systematic patterns emerged: very short runs of 1–3 trials were above-chance but declined toward chance by four trials; runs of 11–19 trials were significantly below chance; and runs beginning at 20 trials showed recurrent spikes at every fifth run length, such as 30, 35, 40, 45, 50. These effects diminished for very long runs (> 80 trials), where overall performance dipped below chance.

Following the reviewers' suggestion, we compared early-trial hit rates between runs that were optionally stopped and runs that continued. Runs stopped after a single trial ($N = 54937$) had a mean hit rate of 0.317, significantly higher than the 0.197 observed for continuing runs ($N = 1{,}601{,}472$), $t(57{,}721.13) = 59.91$, $p < .001$, Cohen's $d = 0.30$, 95% CI [0.29, 0.31]. Runs stopped after 2 trials ($N = 15{,}871$) showed a significant gap (0.287 vs. 0.198, $d = 0.32$, 95% CI [0.30, 0.33]). Runs stopped after 3 trials ($N = 17{,}840$) showed a hit rate of 0.244 vs. 0.198 for continuing runs ($d = 0.20$, 95% CI [0.19, 0.22]). These differences are modest and consistent with the interpretation that optional-stopping behavior selectively retains lucky early outcomes, although they are also consistent with genuine early-trial effects.

Figure 3

Ribbon Plot of Hit Rate, Monte Carlo null envelope



Note. Blue circles indicate empirical outside 95% null envelope.

Figure 3 displays the empirical mean hit-rate curve alongside the 95% null envelope from the Monte Carlo simulation. Several features of the run-length pattern fall within the null envelope and are therefore adequately explained by optional stopping under chance-level performance: the below-chance dip across run lengths 11–19, much of the variability across run lengths 30–100, and many of the round-number points. In contrast, the elevations at run lengths 1, 2, and 20 lie well above the null envelope, and several below-chance points in the 65–97 range fall below the lower envelope. These deviations are not reducible to optional stopping alone and indicate that some features of the empirical curve, particularly the early-trial elevation and the run-length-20 peak, require additional explanation.

To further test whether round-number patterns reflect selective stopping based on performance, we compared cumulative hit rates at each unambiguous round-number stopping point (lengths 30 through 90 in increments of 5) between runs that ended at that length and runs that continued past it (Supplemental Table 10). Run lengths 5, 10, and 25 could not be tested because they coincide with planned run lengths and cannot be classified unambiguously; run length 95 was excluded due to insufficient runs ending at that length. Of the 13 comparisons tested, 8 were significant after FDR correction. Effects were strongest at run lengths 30 and 65 (Cohen's $d = 0.69$ and 0.80, respectively) and attenuated at higher run lengths (no significant effects at run lengths ≥ 80). Across significant comparisons, runs that ended at the round number had cumulative hit rates 2–5 percentage points higher than runs continuing past, consistent with users selectively stopping after favorable performance.

Predictors of Top Performers' Success

An overview of the four preregistered logistic multiple regression models applied to the post-hoc Top Performer subgroup is displayed in Table 4. Two outcomes were modeled (first trial; all trials), each with two model specifications: main effects only, and main effects plus the preregistered interactions.

Table 4

Trial-Level Analysis: Overview of Predictors Across Top Performer Models

Predictor	First Trial: Main Effects	First Trial: Main + Interactions	All Trials: Main Effects	All Trials: Main + Interactions
Intercept	2.71 **	0.39	1.40 ***	1.54
Belief in Psi	1.02 (.004)	1.72	0.92 (-.014) ***	0.93
Precog	0.98 (-.005)	3.67	1.04 (.006) ***	1.07
Meditation	0.91 (-.022)	2.78	1.02 (.004) ***	1.09
Total Trials	1.00 (.000195)	0.89	1.00 (-.000044) ***	1.00
Opt Stop Ratio	3.45 (.292) ***	3.37 ***	2.01 (.113) **	3.85 ***
Belief × Precog	–	0.71	–	0.98
Belief × Meditation	–	0.77	–	0.98
Precog × Meditation	–	0.49	–	0.92
Belief × Total Trials	–	1.00	–	1.00
Precog × Total Trials	–	1.00	–	1.00 ***
Meditation × Total Trials	–	1.00	–	1.00
Belief × Precog × Meditation	–	1.18	–	1.02
Belief × Total Trials × Opt Stop	–	1.00	–	1.00 ***
Precog × Total Trials × Opt Stop	–	1.00	–	1.00 ***
Meditation × Total Trials × Opt Stop	–	1.00	–	1.00 ***

Note. The values are odds ratio (OR), delta probability (*Δp*; the change in probability for a one-unit change in the predictor; therefore intercepts and interactions do not have *Δp*; *Δp* values are also omitted in models with interaction terms because they reflect linear approximations that may be unreliable in those models, see OR for effect direction), and *p*-value levels, with FDR correction (Benjamini–Hochberg) applied to non-intercept predictors within each model. Em dashes (—) indicate that a term is not included in that model. The All Trials models assess all trials from each subject regardless of run structure. The First Trial models assess just the first trial (implicitly, from the first run) submitted by each user. Asterisks indicate the traditional significance thresholds (.05, .01, .001) after adjusting for FDR. Given the very large sample size, significance may be reached for effects that are substantively small; see *Δp* column for effect magnitudes.

Top Performers First Trial

Two logistic models were run for the top performers' first trials only analyses: (1) main effects only (Supplemental Table 5) and (2) main effects plus the preregistered interactions, including three-way terms (Supplemental Table 6).

In all both models, optional stopping (OptStop) was the only significant predictor (besides the intercept); no Psi-type, behavioral, or interaction term reached significance after FDR correction in the preregistered model. In the main-effects model, OptStop was associated with a Δp of 0.29, indicating that a user with 100% optionally stopped runs had, on average, a 0.29 higher probability of hitting on their first trial than a user with 0% optionally stopped runs.

Top Performers All Trials

Two models were conducted for the top performers' all trials analyses: (1) main effects only (Supplemental Table 7), and (2) the preregistered interaction model including three-way terms (Supplemental Table 8).

In the main-effects model, all five predictors were significant after FDR correction, although effect magnitudes were small ($|\Delta p| \leq .014$ for all predictors except OptStop, which had $\Delta p = 0.11$). In the preregistered interaction model (Supplemental Table 8), only the main effect of optional stopping reached significance; main effects of Belief, Precog, Meditation, and Total Trials were no longer significant once preregistered interactions were included. Among interactions, three of the preregistered three-way terms reached significance: Belief × Total Trials × OptStop, Precog × Total Trials × OptStop, and Meditation × Total Trials × OptStop, along with the Precog × Total Trials two-way interaction. As noted in Methods, however, statistical significance at this scale does not imply substantive importance; effect magnitudes were generally small.

In summary, across both first-trial and all-trial analyses, optional stopping was the strongest predictor of success among top performers. In the main effects only models for all trials, Belief, Precog, Meditation, and Total Trials each showed small but statistically reliable associations with hit probability ($|\Delta p| \leq .014$). However, when the preregistered interactions were added, these main effects were no longer significant, while three-way interactions among each Psi-type variable, Total Trials, and Optional Stopping reached significance. This pattern suggests that the influence of belief, prior precognitive experience, and meditation on top performers' performance depends jointly on cumulative task experience and stopping behavior, rather than operating as independent traits. Notably, the main-effect Δp for OptStop in the preregistered all-trials interaction model was extreme ($\Delta p \approx 2.97$), reflecting limitations of the linear approximation when interaction terms are included; the odds ratio (OR = 3.85, $p < .001$) provides a more interpretable estimate of OptStop's role in this specification.

Discussion

This preregistered study examined over 25 million trials from the Quick Remote Viewing task to investigate how psi performance unfolds across run lengths, stopping behaviors, and individual differences. Results reveal that apparent performance fluctuations were not uniform but emerged intermittently, shaped by pacing, engagement, and stopping strategies. From a conservative standpoint, these dynamics may largely reflect known response biases, such as optional stopping. If psi effects exist, they are likely subtle and embedded within these behavioral patterns.

Patterns of Performance in All Participants

On the surface, participants appeared to "know" when to stop, with elevated hit rates after one or two trials and recurring peaks at every fifth run length beginning at 20 (e.g., 30, 35, 40, 45, 50) indicating that performance interacted with self-paced stopping. However, much of this apparent prescience is reproducible from stopping behavior alone. Runs ending at the planned lengths (5, 10, 25, 100) conformed to chance, whereas runs at other lengths, which by definition reflect within-run optional stopping, showed systematic above- and below-chance fluctuations that faded in very long runs (>80 trials). These dynamics could result from expectancy or learning effects but also resemble classic reports of unstable or "psi missing" (Kennedy, 2003). Monte Carlo simulations imposing only optional-stopping rules clarify how much of this structure is reproducible as a behavioral artifact. The 11–19 trough and much of the round-number variability fell within the simulated null envelope, suggesting these patterns are adequately explained by optional stopping under chance-level performance. However, the elevations at run lengths 1, 2, and 20 exceeded the null envelope, indicating that these specific features cannot be reduced to stopping behavior alone. This elevation pattern resembles classic reports of unstable or front-loaded performance in the psi literature (Kennedy, 2003), though we cannot rule out residual stopping-related artifacts at these specific run lengths in this particular study.

At the participant level, variability dominated: three below-chance stretches (4, 11–19, 65–97 trials) alternated with short bursts above chance (1–3 trials, round numbers). Rather than field-wide decline, such patterns likely reflect context-dependent fluctuations driven by participant behavior and task dynamics (Tressoldi & Storm, 2024b). Notably, this finding pertains to study-level effect sizes aggregated across decades of research. These results may differ from the within-individual fluctuations observed here, which reflect performance dynamics at the participant level. Instead, these results suggest that participant-level variability and task context may be primary drivers (cf. Tressoldi & Storm, 2024b, study-level), reframing forced-choice psi tasks as behaviorally dynamic systems where procedure and stopping strategies shape observed outcomes.

Individual Differences and Top Performers

No individual met preregistered significance after multiple-comparison correction, yet in a pool of >46,000 users, even rare, stable deviations may be meaningful. Conservative false-discovery procedures minimize spurious results but also obscure potential "black-swan" performers. Exploratory analyses identified a small high-performer subgroup whose behavior merits confirmation in preregistered, fixed-length retests. A two-stage design, broad screening followed by focused retesting, could better evaluate stability across sessions.

Among these Top Performers, optional stopping was the strongest predictor of success, particularly for first trials (Δp = 0.29 for first trials), with other predictors contributing only marginally ($|\Delta p| \leq .014$). In the interaction model for all trials, the pattern was qualitatively similar: stopping behavior again carried the largest effect, with smaller contributions from belief, meditation, and prior precognitive experience emerging only through their three-way interactions with Total Trials and Optional Stopping. These findings suggest that situational and behavioral factors, more than enduring traits, shape observed outcomes.

An alternative interpretation of the three-way interactions involving Total Trials and Optional Stopping (Belief × Total Trials × OptStop, Precog × Total Trials × OptStop, Meditation × Total Trials × OptStop; see Supplemental Table 8) is worth acknowledging. Participants who complete very large numbers of trials have ample opportunity to refine informal stopping rules, such as stopping after perceived "hot" streaks, at psychologically salient milestones, or when early performance seems favorable. These refinements may interact with trait-level differences in belief, meditative practice, and prior precognitive experience. The result may be inflated hit rates through outcome-dependent stopping rather than enhanced psi accuracy that varies as a function of those traits. This perspective is compatible with the broader behavioral-dynamics framing of this study and does not require assuming that psi ability changes as a function of experience or as a function of belief, meditation, or prior precognitive experience.

Methodological Implications

Psi research traditionally relies on pooled binomial tests assuming trial independence (Jahn & Dunne, 1987; Utts, 1991). In the QRV context, such aggregation can mask sequential clustering and behavioral dynamics. Once deviations appear, independence assumptions weaken, and apparent effects may arise from expectancy, fatigue, or stopping biases. Alternatively, if psi exists, its operation may be intermittent and state-dependent, intertwined with ordinary variability. Mixed-effects and repeated-measures models (Storm et al., 2010; Tressoldi & Storm,

2024b) are better suited to capture these temporal fluctuations and individual differences than fixed-effect binomial tests.

Because massive datasets yield minuscule *p*-values even for trivial deviations (Wagenmakers et al., 2012), reporting effect sizes, confidence intervals, and Bayesian estimates are essential. Preregistration further constrains analytic flexibility yet aligns psi research with open-science norms (Nosek et al., 2018).

Implications for Parapsychology

At a broader level, this large-scale QRV study underscores both the opportunities and challenges inherent in modern, open-ended online psi testing. Compared with traditional fixed-length laboratory tasks, at-will designs introduce substantial methodological complexity, particularly through optional stopping and unstandardized participant behavior, highlighting the need for refined analytic approaches rather than simply replicating classic forced-choice methods. The influence of stopping behavior and participant variability underscores the limits of examining only aggregated trial data and supports the integration of participant-level modeling as a standard practice (Tressoldi & Storm, 2024b). Rather than weakening evidential standards, this approach complements pooled binomial tests with analyses that investigate the mechanisms driving anomalous cognition. Framing psi research in terms of both proof and process complements pooled binomial tests with participant-level models that investigate mechanisms, while highlighting challenges inherent in open, online testing. This integrative perspective also aligns with contemporary movements in consciousness studies that emphasize linking first-person experience, interactive dynamics, and third-person measurement (Cardeña, 2018; Kripal, 2019).

Interpreting Psi Dynamics

The QRV data are compatible with multiple interpretations. The dominant role of optional stopping behavior could reflect a purely psychological bias (expectancy, fatigue, or preferences for round-number runs), or it could reflect subtle psi-related intuitive or embodied cues influencing the moment of decision (i.e., the decision to stop as itself a form of psi). The present data cannot distinguish these possibilities. This is a question best answered with preregistered designs combining phenomenological and physiological measures.

A more skeptical reading of the same pattern is also possible. The absence of effects in runs of 100 trials, the only run length that can be unambiguously identified as a completed, fixed-

length attempt (see Methods, Task Description), is itself a substantive finding. It is compatible with at least three interpretations: that psi manifests primarily under novelty or uncertainty and does not survive pre-committed task structures (Kennedy, 2001, 2003); that psi operates by influencing *when* participants act rather than *what* outcome occurs, such that fixed-length runs eliminate the relevant degree of freedom (decision augmentation theory; May et al., 1995); or that apparent effects in optionally stopped runs reflect stopping-related artifacts rather than anomalous cognition. These interpretations converge on a common methodological recommendation: fixed-length, preregistered designs are essential for distinguishing candidate psi signals from behavioral artifacts.

Regression analyses also indicate a front-loaded effect, with higher probabilities of success on the very first trial. Under a psi interpretation, this would be consistent with both a genuine peak early in performance (Kennedy, 2003; Mossbridge & Radin, 2018) and with selective retention of early hits into the Top Performer subgroup. Such episodic patterns mirror findings from spontaneous cases and Ganzfeld experiments, where psi appears briefly and context-dependently (Wahbeh et al., 2018; Tressoldi & Storm, 2024a). Descriptive hit rates across nested trial subsets (Table 2) further illustrate this front-loaded pattern among Top Performers, where hit rates declined monotonically from approximately 50% on the first trial to chance levels across all trials. However, this decline is also consistent with selective retention: because Top Performers were defined by overall above-chance performance, users whose early trials were hits are over-represented, particularly at shorter subsets. The present data cannot distinguish a genuine front-loaded psi effect from this selection dynamic.

Trait-level predictors, such as belief, meditation, or prior precognitive experience, were weak. This may reflect limitations of the brief self-report measures used at registration, or it may indicate that stable individual differences play a smaller role than situational and behavioral factors in this paradigm. Performance declined with heavy repetition, consistent with fatigue or boredom, but also with proposals that psi is transient or self-limiting (Kennedy, 2001). Future work should distinguish between ordinary motivational decline and possible self-limiting psi dynamics.

Limitations

The QRV dataset has several limitations to be considered when reviewing the results. Participants did not provide demographic information, such as age, gender, or race, which prevented examining how these variables might moderate psi performance. Because usernames were self-generated, some individuals may have registered multiple times under different aliases.

Although the dataset's massive scale (>25 million trials) likely minimizes the effects of duplication on pooled outcomes, it reduces certainty in participant-level and repeated-measures analyses. As noted in the Methods, the dataset records the number of trials completed in each run but not the participant's chosen planned length. This prevents clean separation of completed planned runs from optionally stopped runs at the 5-, 10-, and 25-trial lengths, and conclusions about performance at these specific lengths should therefore be interpreted with that constraint in mind. Some analytic choices, such as the post-hoc redefinition of top performers and the use of log-transformed run counts, were not preregistered but were adopted to maintain feasibility and interpretability. The preregistered fixed-effect logistic regression approach does not formally account for the nested structure of trials within users. Mixed-effect models with random intercepts for user and appropriately scaled predictors would be a more appropriate alternative for confirmatory replications.

Brief registration questionnaires captured handedness, creativity, RV training, and belief/experience measures, but these were single-item indicators that may not have been able to detect subtle relationships. Self-selection likely produced a sample skewed towards believers, limiting variance and reducing power for trait–performance correlations. The group comparison in Table 2, though not preregistered, was included for descriptive transparency and showed that small numerical differences between Top and Not Top performers were unlikely to explain the preregistered findings. Future studies could include fuller multi-item psychological scales, demographic variables, and targeted modeling to clarify moderators such as handedness, creativity, and RV training. Finally, uncontrolled device, browser, and environmental differences may have introduced noise or subtle biases not accounted for in the current analyses.

Future Directions

These findings highlight key directions for future research. Mixed-effects and multilevel models remain essential for accommodating individual variability and session dynamics while retaining sensitivity to meaningful effects (Tressoldi & Storm, 2024b). Repeated-measures designs can help distinguish stable individual differences from noise and clarify how psi unfolds over time.

Going beyond aggregate hit rates, trial-level analyses of effect sizes and temporal structure will be critical for revealing subtle patterns obscured in pooled data. Addressing biases from optional stopping may require sequential or adaptive statistical approaches that legitimately accommodate ongoing data collection. Combining these methods with signal detection theory (SDT) can separate perceptual sensitivity from decision criteria while remaining valid under optional stopping. SDT offers a powerful framework for assessing whether participants extract

information even when guessing incorrectly, helping determine whether anomalous cognition reflects genuine information transfer or shifts in decision-making (Storm et al., 2010; Utts, 1991; Anderson, 2015; Román et al., 2022).

Monte Carlo simulations also provide valuable baselines for modeling complex stopping behaviors, sequential dependencies, and noise influences that violate trial independence (Varvoglis et al., 2019). Measures such as autocorrelation can reveal whether hits and misses cluster over time, clarifying whether effects arise from session-level dynamics rather than isolated trials (Fischer & Whitney, 2014).

More elaborate temporal analyses remain for future work. Sequential Bayesian updating could track how evidence for above-chance performance accumulates trial by trial within each user, without the Type I inflation of repeated frequentist testing. Cross-session stability analyses could test whether users who perform above chance in one session continue to do so in subsequent sessions, distinguishing stable individual differences from session-level noise.

Future studies should incorporate demographic, cultural, and biological moderators (e.g. race, sex, handedness, social role) and psychological factors such as belief systems and personality traits to test whether these systematically influence psi expression (Storm et al., 2010). Expanding recruitment beyond psi-interested populations and integrating these contextual layers will move the field toward a more relational understanding of psi as a human, rather than purely statistical, phenomenon. Continued methodological refinement and broader collaboration will be essential for translating insights from large-scale datasets into the next generation of study designs.

Conclusions

Analyses of over 25 million QRV trials revealed that performance dynamics reflect complex interactions among stopping behavior, task context, and individual variability. Group-level results were driven by optional stopping, while exploratory top-performer analyses suggested a context-dependent interplay among belief, experience, and meditation that influenced performance only through their joint interactions with task experience and stopping behavior, rather than as independent predictors. These findings underscore that psi, if present, may emerge episodically and in interaction with behavioral states, calling for preregistered, fixed-length, participant-level designs capable of distinguishing subtle signals from the noise of human behavior.

Acknowledgments

The authors thank the Institute of Noetic Sciences (IONS) for hosting and maintaining the GotPsi platform and for providing administrative and technical support over the years of data collection. We are especially grateful to the thousands of participants who engaged with the online task and contributed to this long-term citizen-science project. The authors also acknowledge the assistance of IONS research staff and volunteers who supported aspects of data management and manuscript preparation.

Disclosures

Author Contributions

Helané Wahbeh conceptualized the study, developed the preregistration and analytic plan in collaboration with Michael Kriegsman, and led the interpretation of results and manuscript preparation.

Michael Kriegsman performed all data cleaning, statistical analyses, and figure generation, and contributed to data interpretation and manuscript editing.

Beth Glick contributed to the drafting and revising of the Introduction and Discussion sections and assisted with manuscript preparation through multiple iterations.

Arnaud Delorme co-managed the GotPsi website and dataset, reviewed the preregistration and analytic framework, and provided substantive feedback on the final manuscript.

Dean Radin created and maintained the GotPsi website, contributed to the long-term data collection, reviewed the preregistration and statistical plan, and provided critical revisions of the manuscript.

AI Use Disclosure

Portions of this manuscript were prepared with the assistance of OpenAI's ChatGPT, which was used to support drafting, summarization, and editorial refinement of text throughout the project. The authors guided all content generation and are solely responsible for the final versions of all analyses and interpretations.

Correspondence concerning this article should be addressed to Helané Wahbeh, ND, MCR, Institute of Noetic Sciences; 7250 Redwood Blvd, Ste 208, Novato, CA 94945-327; hwahbeh@noetic.org

Conflict of Interest

The authors declare no conflicts of interest.

Funding

This research received no external funding.

Open Data

Data processing code and analysis scripts will be shared upon publication (with privacy safeguards).

References

Anderson, N. D. (2015). Teaching signal detection theory with pseudoscience. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00762

Baptista, J., Derakhshani, M., & Tressoldi, P. E. (2015). Explicit anomalous cognition: A review of the best evidence in Ganzfeld, forced choice, remote viewing and dream studies. In E. Cardeña, J. Palmer, & D. Marcusson-Clavertz (Eds.), *Parapsychology: A handbook for the 21st century* (pp. 192–214). McFarland & Company, Inc., Publishers.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Cardeña, E. (2018). The experimental evidence for parapsychological phenomena: A review. *American Psychologist, 73*(5), 663–677. https://doi.org/10.1037/amp0000236

Dalkvist, J., Mossbridge, J., & Westerlund, J. (2014). How to remove the influence of expectation bias in presentiment and similar experiments: A recommended strategy. *Journal of Parapsychology, 7*8(1), 80–97.

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience, 17*(5), 738–743. https://doi.org/10.1038/nn.3689

Hyman, R. (1995). Evaluation of program on "anomalous mental phenomena." *Journal of Parapsychology, 59*, 321–351.

Jahn, R. G., & Dunne, B. J. (1987). *Margins of reality: The role of consciousness in the physical world*. Jovanovich.

Kennedy, J. (2001). Why is PSY so elusive? A review and proposed model. *The Journal of Parapsychology, 65*, 219–246.

Kennedy, J. (2003). The capricious, actively evasive, unsustainable nature of psi: A summary and hypotheses. *The Journal of Parapsychology, 67*(1), 53–75.

Kennedy, J. (2013). Methodology for confirmatory experiments on physiological measures of precognitive anticipation. *The Journal of Parapsychology, 77*(2), 237.

Kennedy, J. (2014). Letter on methodology for presentiment studies. *Journal of Parapsychology, 78*(2), 273–274.

Kennedy, J. (2016). Is the methodological revolution in psychology over or just beginning. *Journal of Parapsychology, 80*(2), 156–168.

Kripal, J. J. (2019). *The flip: Epiphanies of mind and the future of knowledge*. Bellevue Literary Press.

May, E. C., Utts, J., & Spottiswoode, S. (1995). Decision Augmentation Theory: Toward a model of anomalous mental phenomena. *Journal of Parapsychology, 59*(3), 195–220.

Mossbridge, J., & Radin, D. (2018). Precognition as a form of prospection: A review of the evidence. *Psychology of Consciousness: Theory, Research, and Practice, 5*(1), 78–93. https://doi.org/10.1037/cns0000121

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences, 115*(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114

Palmer, J. (1985). *An evaluative report on the current status of parapsychology*. U. S. Army Research Institute for the Behavioral and Social Sciences.

R Development Core Team, R. (2023). R: A language and environment for statistical computing [Computer software]. *R Foundation for Statistical Computing*. http://www.R-project.org/

Rhine, J. B. (1934). *Extra-sensory perception*. Boston Society for Psychic Research.

Román, C. A. F., DeLuca, J., Yao, B., Genova, H. M., & Wylie, G. R. (2022). Signal detection theory as a novel tool to understand cognitive fatigue in individuals with multiple sclerosis. *Frontiers in Behavioral Neuroscience, 16,* 828566. https://doi.org/10.3389/fnbeh.2022.828566

Storm, L., & Tressoldi, P. E. (2020). Meta-analysis of free-response studies 2009-2018: Assessing the noise-reduction model ten years on. *Journal of the Society for Psychical Research, 84*(4), 193–219.

Storm, L., Tressoldi, P. E., & Di Risio, L. (2010). Meta-analysis of free-response studies, 1992–2008: Assessing the noise reduction model in parapsychology. *Psychological Bulletin, 136*(4), 471.

Storm, L., Tressoldi, P. E., & Di Risio, L. (2012). Meta-analysis of ESP studies, 1987-2010: Assessing the success of the forced-choice design in parapsychology. *Journal of Parapsychology, 76*(2), 243–273.

Tressoldi, P. E., & Storm, L. (2024a). Stage 2 Registered Report: Anomalous perception in a Ganzfeld condition-A meta-analysis of more than 40 years investigation [version 4; peer review: 2 approved, 1 not approved]. *F1000Research, 10*, 234. https://doi.org/10.12688/f1000research.51746.4

Tressoldi, P. E., & Storm, L. (2024b). The myth of the decline effect in psi research: The empirical evidence. *Journal of Scientific Exploration, 38*(3), 461–465. https://doi.org/10.31275/20243313

Utts, J. (1991). Replication and meta-analysis in parapsychology. *Statistical Science, 6*(4), 363–378.

Utts, J. (1996). An assessment of the evidence for psychic functioning. *Journal of Scientific Exploration, 10*(1), 3–30.

Varvoglis, M., Bancel, P.A., Bailly, J.-P., Boban, J., & Ahmed, D.S. (2019). The Selfield: Optimizing precognition research. *Journal of Parapsychology, 83*(1), 13–24. https://doi.org/10.30891/jopar.2019.01.02

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H.L.J., & Kievit, R.A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science, 7*(6), 632–638. https://doi.org/10.1177/1745691612463078

Wahbeh, H., Radin, D., Delorme, A., & Kriegsman, M. (2025). Probing top performers in a forced choice precognition task [Preregistered study]. *KPU Study Registry*. https://www.koestler-parapsychology.psy.ed.ac.uk/Documents/KPU_Registry_1094.pdf

Wahbeh, H., Radin, D., Mossbridge, J., Vieten, C., & Delorme, A. (2018). Exceptional experiences reported by scientists and engineers. *Explore: The Journal of Science and Healing, 14*(5), 329–341. https://doi.org/10.1016/j.explore.2018.05.002

Zusammenfassung

Untersuchung von Top-Performern bei einer Forced-Choice-Hellsehaufgabe

Diese präregistrierte Studie analysierte mehr als 25 Millionen Durchgänge einer webbasierten „Forced-Choice"-Remote-Viewing-Aufgabe, um Muster der hellseherischen Leistung bei allen Teilnehmern sowie bei einer Untergruppe der besten Teilnehmer zu untersuchen. Auf aggregierter Ebene entsprachen Durchgänge, die nach den vier geplanten Längen (5, 10, 25 oder 100 Trials) endeten, den Zufallserwartungen. Im Gegensatz dazu zeigten Durchgänge mit optionalem Abbruch („optional stopping") systematische Schwankungen: Kurze Durchgänge (1–3 Versuche) lagen über dem Zufallsniveau, bevor sie abfielen; Durchgänge mit einer Länge von 11–19 Versuchen lagen unter dem Zufallsniveau; und Durchgänge ab einer Länge von 20 Versuchen zeigten wiederkehrende Spitzenwerte über dem Zufallsniveau bei jeder fünften Durchgangslänge (z. B. 30, 35, 40, 45, 50), die bei mehr als 80 Versuchsdurchgängen abnahmen. Eine an die empirische Abbruchverteilung angepasste Monte-Carlo-Simulation verdeutlichte, inwieweit diese Muster allein durch ein optionales Abbruchverhalten reproduziert werden konnten, wobei ein Großteil des Musters der Anzahl der Versuchsdurchgänge – einschließlich des Tiefpunkts bei 11–19 und der Variabilität bei runden Zahlen – innerhalb der simulierten Nullhüllkurve lag. In explorativen Analysen der Top Performer – post-hoc definiert als jene 1.235 Nutzer (2,64 %), die das unkorrigierte Zufallsniveau übertrafen, nachdem kein Nutzer das präregistrierte, FDR-korrigierte Kriterium erfüllt hatte – wurden der Glauben an Psi, frühere präkognitive Erfahrungen, Meditation, die Gesamtzahl der Versuchsdurchgänge sowie der

optionale Abbruch als Prädiktoren untersucht. Der optionale Abbruch war der Prädiktor, der sowohl bei den ersten Versuchen als auch über alle Versuche hinweg am konsistentesten mit Treffern assoziiert war; zudem zeigte er Interaktionseffekte mit den Glaubensüberzeugungen, früheren präkognitiven Erfahrungen und Meditation in Verbindung mit der kumulativen Erfahrung in der Aufgabe. Die Effektstärken waren gering (Δp und Cohens d lagen für die meisten Prädiktoren nahe null), und die Ergebnisse werden als explorativ interpretiert. Die Befunde legen nahe, dass die Ergebnisse auf Gruppenebene primär den optionalen Abbruch und damit verbundene Verhaltensdynamiken widerspiegeln, während die Analysen der Top Performer differenziertere, wenn auch schwache, kontextabhängige Zusammenhänge zwischen Glaubensüberzeugungen, Erfahrung und Verhalten aufzeigen. Diese Befunde unterstreichen die methodischen Herausforderungen groß angelegter, offener Online-Tests sowie den Wert präregistrierter Ansätze auf Teilnehmerebene in Kombination mit Benchmark-Simulationen, um Verhaltensartefakte von potenziellen Psi-Signalen zu unterscheiden.

Schlüsselbegriffe: Hellsehen, Forced-Choice-Aufgabe, Optional Stopping, Top-Performer, individuelle Unterschiede, Präregistrierung

Journal of Anomalistics
Zeitschrift für Anomalistik

www.anomalistik.de www.igpp.de

Development of a Dual-Mode Application for Psi Research

Brian Laythe [a,*], James Houran [b,c]

Abstract – We present the development and pilot evaluation of the Immersive Psi Test (IPT), a dual mode application for research on anomalous cognition that integrates forced choice measurement with AI supported narrative immersion to enhance ecological engagement. The IPT combines five image-based forced choice trials ("explicit psi") with ten covert narrative decisions ("implicit psi"), using normed "enchanted" (high emotion and numinosity) and "disenchanted" (low emotion and numinosity) photographs to influence aesthetic-transformative experience. Enchantment Manipulation checks confirmed strong valence separation ($d = 3.63$) and broadly acceptable randomization, with only minor deviations in one trial. In a convenience-sample pilot study ($N = 126$), explicit psi performance with image selections did not exceed chance, whereas implicit psi via narrative choices in the enchanted condition produced a very small but statistically significant deviation above chance ($p = .016$; $h = 0.09$). Measures of transliminality and encounter proneness correlated as expected ($\rho = .52$) but neither predicted psi outcomes in our sample. These findings demonstrate the operational feasibility of the IPT framework and provide potential evidence of condition-specific implicit psi effects. Reliance on a convenience sample limits generalizability, so the outcomes should be regarded as a proof-of-concept versus confirmatory evidence. Future research should employ preregistered designs, stratified sampling, and enhanced randomization controls to determine whether observed effects reflect genuine anomalies or methodological artifacts.

Keywords: immersive testing, instrumentation, liminality, narrative engagement, psi

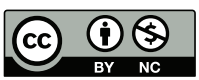
(a) Institute for the Study of Religious and Anomalous Experience, Charlestown, Indiana, USA

(b) Integrated Knowledge Systems, Meridian, Idaho, USA

(c) Research Group on Economic Behavior and Subjective Well-Being, Rey Juan Carlos University, Madrid, Spain

[*] Corresponding author: Brian Laythe

blaythe@israenet.org



The publication of new research protocols, conceptual frameworks, and methodological recommendations is increasingly recognized as valid and valuable scholarly contributions. This reflects a broader movement within open science to prioritize transparency, reproducibility, and cumulative theory-building (Bandrowski et al., 2023; Mello et al., 2020; Springer Nature Research Communities, 2021). Such innovation is particularly salient in parapsychology, where the phenomena under study – including ostensible nonlocal perception, mind-matter interactions, and firsthand accounts of anomalous experiences – intersect with foundational questions about the nature and limits of consciousness (Wahbeh et al., 2022). Numerous authors have therefore advanced guidelines and frameworks for activities in this context (e.g., Baker & O'Keeffe, 2007; Houran, Maraldi et al., 2025; Lange, 2017; Laythe et al., 2021; Thomas et al., 2026; Zemel & Wahbeh, 2025), which emphasize the importance of structured, ethical, and reflexive approaches. Moreover, parapsychology has long served as a proving ground for procedural controls, analytical rigor, and sophisticated technologies that have influenced research standards in broader scientific domains (e.g., Bierman et al., 2016; Kennedy, 2016; Morris, 2001; Pooley, 2025; Sheldrake, 1998; Watt & Nagtegaal, 2004).

We aim to extend this tradition with a novel framework to investigate *psi* – i.e., anomalous process(es) of apparent information acquisition or influence that ostensibly resist description by known physical or biological mechanisms (for a review, see Cardeña, 2018). Psi's controversial status aside, its potential to reshape our understanding of perception and consciousness makes it a legitimate target for scientific inquiry. Several mainstream journals across different disciplines have published parapsychological research (e.g., Bem et al., 2015; Escolà-Gascón et al., 2022; Kekecs et al., 2023; Rabeyron & Watt, 2010; Radin, 2025; Walleczek et al., 2025) although investigating psi as a testable hypothesis is not synonymous with asserting its ontological reality (Schooler et al., 2018). Careful, methodologically diverse studies are therefore essential to determine whether positive findings reflect genuine anomalies or arise from experimental or analytical artifacts. To that end, we developed a dual-mode application that combines the precision of objective scoring with the experiential richness of immersive, free-response environments that presumably support psi. This scalable design is easy to replicate across settings and populations and advances psi methodology in line with open-science principles of transparency, reproducibility, and theoretical integration.

The primary contribution of this work is the development and pilot validation of a research protocol rather than hypothesis testing. Accordingly, we pursue three objectives: (1) articulate the theoretical and empirical rationale for a hybrid "forced-choice × immersive" psi application; (2) detail the system's technical architecture – covering design parameters, scoring algorithms, and implementation procedures – in a way that ensures full transparency and enables reliable

replication and adaptation in future studies; and (3) present pilot data evaluating the application's operational feasibility and producing exploratory psi estimates for use in subsequent power calculations. These results also offer preliminary insights into the strengths and limitations of our hybrid model, thereby informing recommendations for its refinement and broader application. The remainder of the paper is therefore structured to address these three objectives in order, guiding the reader from theoretical rationale to technical specification and finally to pilot evaluation.

1. Background: Rationale for a Dual-Mode Application for Psi Research

Forced-choice paradigms have long been favored for their clear scoring rules and the deliberately high perceptual contrast built into their target sets – most famously the Zener cards, which were explicitly designed to minimize ambiguity among symbols. Their straightforward statistics and strong cross-laboratory replicability further enhance experimental credibility and help to address field-level skepticism (Radin, 1997; Storm et al., 2012). Meta analytic summaries likewise show small but reliable forced choice effects: Honorton and Ferrari (1989) reported a combined $z = 11.41$ (trial based mean $ES \approx .02$), with larger effects when studies used selected participants, individual testing, trial by trial feedback, and short response to target intervals; Storm and Tressoldi (2023) similarly found a small average effect (a few hundredths of a standardized effect) and concluded that forced choice designs produce consistent but modest effects relative to free response protocols such as used in ganzfeld (i.e., sensory deprivation) and remote viewing (RV) studies.

Free-response protocols trade the objectivity of forced-choice designs for greater ecological validity by immersing participants in "high-entropy" stimuli – complex, distinctive, surprising, or information-rich targets – using the term in a psychological sense rather than its thermodynamic one. In the ganzfeld, for instance, a "receiver" in sensory reduction describes impressions a "sender" attempts to transmit; in RV a percipient describes a hidden or distant target and independent judges match descriptions to targets. Ganzfeld programs with rigorous controls – automated selection, blind judging, and detailed logging – reported above chance performance and suggested that dynamic targets (video clips) increase hit rates compared with static images (Honorton et al., 1990). Subsequent reviews and meta analyses qualified those findings by highlighting heterogeneity across labs and sensitivity to procedural details (Hyman, 1985; Milton & Wiseman, 1999). Recent work has moved toward standardized free response evaluation to support efficiency, consistency, and objectivity, including preliminary artificial intelligence (AI) based scoring methods (e.g., Mossbridge et al., 2025), but challenges of judging consistency and interpretive bias persist (Hyman & Honorton, 1986; Milton & Wiseman, 1999).

Stimulus salience and affective relevance also emerge repeatedly as plausible moderators across protocols. Bem (2011) reported small but significant retroactive effects that were larger for affectively salient targets, implying that motivational relevance or attention capturing content may amplify reported effects. Many large, preregistered replications; however, failed to reproduce Bem's findings, and critics have pointed to analytic flexibility, publication bias, and confounds as alternative accounts (Galak et al., 2012; Ritchie et al., 2012; Simmons et al., 2011; Wagenmakers et al., 2011). Meta analyses confirm that affective stimuli tend to show larger effects but also show that differential attention, arousal, expectancy, and post hoc analytic choices can mimic psi like outcomes. Thus, affective or numinous imagery should be treated as a testable moderator, not a confirmed mechanism.

Both forced-choice and free-response literatures converge on concrete methodological safeguards: predefine and balance target sets (dynamic vs. static; high vs. low affective salience), automate selection and judging where feasible, preregister analytic plans, and pursue high powered direct replications to separate true target dependent modulation from artifacts such as selective reporting or attention/arousal confounds (Bem, 2011; Honorton et al., 1990; Hyman, 1985). Theoretically, if target features reliably influence psi effects, models must allow receiver–target interactions in which dynamic complexity and affective relevance alter attentional, motivational, or intersubjective coupling; until stringent, preregistered replications demonstrate such modulation, ordinary psychological mechanisms remain viable explanations (Hyman, 1985; Milton & Wiseman, 1999). Recent RV work reinforces this pragmatic stance. Specifically, Tressoldi and Katz (2023) reported a robust average effect and noted that richer, higher bandwidth targets – dynamic video clips and complex scenes – are associated with larger effect sizes; earlier work similarly reported greater perceivability for dynamic versus static targets (Lantz et al., 1994).

Evidence for the influence of numinous or high valence targets is likewise suggestive but sparse: only a few studies have examined target interest or contextual salience (Katz et al., 2021), and Krippner et al. (2019) tested light versus dark conditions while probing whether targets construed as more numinous produced different accuracy patterns, reporting results that justify systematic valence investigations. Tressoldi and Katz (2023) emphasized that procedural variables – interviewer presence, outbound agents, participant selection, and scoring/judging – often correlate with target properties and can inflate or masquerade as target effects. Consistent with this broader pattern, Watt's (1996) three-experiment series found no reliable advantage for emotionally charged targets in forced-choice ESP tasks: emotional versus neutral images produced equivalent scoring once response biases were controlled, and exploratory analyses using participants' own emotionality ratings likewise showed minimal differentiation.

Complementing these findings, Parker et al.'s (1998) ganzfeld work suggests that spontaneous emotional shifts or "aha" reactions during the session – not the emotional content of the targets per se – tend to accompany successful trials. Perhaps the most parsimonious conclusion is that effective psi targets are vivid, distinctive, and moderately complex – whereas the idea that numinous or high-valence content alone boosts psi performance remains plausible but unsubstantiated. Resolving this question will require preregistered, well-powered manipulations of dynamism and valence with strict controls for judging, monitoring, and participant selection (Katz et al., 2021; Tressoldi & Katz, 2023).

Ideal forced-choice targets are therefore clearly discrete items that map directly onto response options – for example, labeled static images, short single-event video clips, predefined object sets, or categorical locations. Emotional or dynamic stimuli should be categorized in advance (e.g., high vs low valence or dynamic vs static) so that responses fall into objective categories, and complex scenes should be reduced to a small set of pre-specified salient features (faces, water, motion) scored as presence/absence items. Considerable evidence shows that visual stimuli – such as certain paintings, photographs, or digital media – can carry intensely evocative qualities and, in some cases, elicit a sense of numinosity or enchantment in viewers (Houran et al., 2026). To clarify, "enchantment" denotes a distinctive mental state marked by dissonance and ontological shock, arising when ordinary waking experience is abruptly disrupted by a profound, meaning-laden awareness that evokes a transcendent sense of connection to an "ultimate reality" (see e.g., Drinkwater et al., 2022). Protocols might further mechanize randomization and recording, test participants individually, and use blinded – preferably automated or AI-assisted – scoring with a preregistered analysis plan to support consistency and objectivity. Studies should be powered for small effects and orthogonally manipulate dynamism and valence while controlling attention and arousal so any performance modulation can be attributed to stimulus–protocol interactions rather than judging bias, analytic flexibility, or selective reporting.

2. Methodology: Development of a Dual-Mode Application for Psi Research

The forced-choice and free-response literatures collectively suggest several practical prescriptions: (1) use forced choice scoring when targets can be discretized into low ambiguity, well balanced alternatives (high quality *k* choice sets or objectively coded feature bins), (2) automate randomization and recording, (3) predefine and preregister the analytic plan, and (4) control examiner, judging, and feedback procedures to remove experimenter and selection confounds. To balance precision with immersive engagement, we therefore created a scalable,

hybrid forced-choice framework that use AI-generated narratives with both visual and thematic contents that adapt to participants' selections. These narratives notably feature photographic images with proven immersive and transformative potential (Houran & Laythe, 2026), seamlessly integrated into a structured, selectable response format.

Advances in large language models (LLMs) – i.e., highly developed AI systems trained to understand and generate human-like text – and adaptive testing environments have thus enabled the simulation of emotionally engaging scenarios or exploratory settings while preserving the methodological control of traditional forced-choice designs. We argue that this fusion generally enhances participant engagement, reduces fatigue (Hyman, 1985), and leverages affective cues that are hypothesized to facilitate anomalous cognition (Targ, 2012), all without sacrificing scoring objectivity or statistical power. Our hybrid model thus merges three core components: (1) the unambiguous response options of forced-choice testing, (2) the experiential depth of immersive storytelling aligned to participants' idiosyncratic preferences in certain respects, and (3) the adaptability and scalability of AI-guided vignettes.

This framework allows participants to engage with a branching narrative modeled on the classic *Choose Your Own Adventure* gamebooks, which present stories in the second person and invite readers to assume the role of the protagonist by making decisions that shape the character's actions and the unfolding plot (Montgomery & Packard, 1979–1998). Each decision point presents intuitive or affective cues followed by five structured options, one of which our system randomly and covertly designates as a "correct hit" (i.e., the psi target). This design retains the advantages of a forced-choice paradigm – such as clear scoring, standardized stimulus presentation, precise control over sensory and symbolic elements – while also offering the cognitive engagement and motivational benefits of a semi-naturalistic, interactive, and aesthetic media experience (cf. Annett et al., 2016; Houran et al., 2026). In fact, it reflects principles of "gamified" platforms that transform everyday interactions into motivating, rewarding, and immersive experiences (Buckley & Doyle, 2016; Jaramillo-Mediavilla et al., 2024; Li et al., 2024).

The present "Immersive Psi Test" (IPT) application extends Laythe and Robert's (2022) original mobile test design and unfolds as a second-person narrative experience lasting ≤ 10-minutes. Participants receive the following instructions and priming for the current protocol:

> Welcome to the "Choose Your Own Paranormal Adventure." We selected you to participate because you scored as a highly sensitive or intuitive person that might have psychic-type abilities. Therefore, we want you to use your "psychic intuition" to complete five simple tasks involving a series of photographs that help to create a mystical story in which you will choose from different "portals" depicted by photographs on the screen. Please use your powerful imagination as you read each question and then use your intuition to

> select a specific photograph as directed. Each time you choose a photograph, you will be "transported" to that location, where you will get to explore two areas via text description, and then you will return to another task involving another photograph selection. It's fun and only takes about 10 minutes to finish. You have been transported to a place where five portals appear. One portal has been secretly chosen by the spirit in the machine. Focus for a moment, breathe, and prepare to select the portal you sense is the target.

As the AI-generated story progresses, the application dynamically adapts subsequent narrative branches based on participants' prior selections, creating an immersive experience while preserving analytic tractability through its forced-choice structure. Note that the narrative language embedded in the instructions (e.g., references to "portals," "transportation," and "the spirit in the machine") serves as a gamified framing device intended to enhance engagement and immersion. This framing encourages participants to rely on intuition and imaginative focus rather than deliberate analytical reasoning; a strategy commonly used in experimental paradigms designed to elicit intuitive responding. Importantly, these narrative elements function as procedural prompts rather than theoretical claims about the mechanism underlying the task, and the wording was presented identically to all participants to maintain standardized conditions. The IPT therefore combines experimental control with an immersive narrative context designed to sustain attention and motivation during the brief testing session.

The following subsections detail the construction and coding aspects of the IPT application, as well as related validation of randomness involving both boot-strap and pilot sample analysis. We collected data on selection accuracy against the AI's hidden sequence and response times (in milliseconds) at each decision node for the psi trials. This combination of behavioral and subjective measures permits rigorous statistical analysis while providing insight into how immersive features impact both putative psi performance and the user experience.

2.1 IPT Procedural Flow

We built the IPT application to support a range of functionalities as needed, including: (1) administration of psychometric measures to screen for inclusion/exclusion criteria or to collect psychological data for covariates of psi performance (e.g. Baptista et al., 2015), (2) an alternative research task for participants not meeting inclusion criteria, and (3) both "explicit" and "implicit" indices of putative psi. Explicit Psi refers to conscious, reportable parapsychological-like experiences, e.g., someone who knowingly claims to instigate mind-matter interactions, intentionally used a psychic technique, or can verbally describe the content and context of a precognitive vision. These experiences are accessible to the person's awareness and can

be directly measured with self-report, structured interviews, or tasks that invite participants to intentionally use or report their claimed ability. In contrast, Implicit Psi denotes ostensible parapsychological effects that occur outside conscious awareness or intention. People do not necessarily notice these effects or label them as "psychic or paranormal." Instead, it manifests as subtle patterns in cognition or behavior (shifted choices, reaction-time differences, or statistical deviations from chance) that may require indirect measures or careful analysis to detect (e.g., Bem, 2011; Radin, 2006).

Standardized image databases are readily available (e.g., Bendall et al., 2025; Kurdi et al., 2017; Lang et al., 1995; Marchewka et al., 2014), but the IPT uses a series of 25 "enchanted" and 25 "disenchanted" images that Houran and Laythe (2026) carefully curated from royalty-free sources and subsequently normed either to exhibit or lack six environmental gestalt effects (cf. Houran et al., 2023) that foster psychological absorption, such as *Affordance* ("This image is full of interesting possibilities that I might like to explore"), *Ambiguity & Threat* ("This image makes me feel anxious or uncomfortable, as if I'm on high alert"), *Atmosphere* ("This image definitely has an overall emotion or mood to it"), *Immersion & Presence* ("This image totally grabs my attention"), *Legibility* ("This image has a layout and details that are easy for me to remember"), and *Memory & Associations* ("This image has colors, symbols, objects, or a sense of time with personal meaning to me"). These collective features aimed to stimulate immersive-aesthetic experiences (Houran et al., 2026) in participants assigned to the IPT's Enchanted-stimuli condition.

Specifically, the Enchanted images comprised five distinct thematic "scenes": (1) *Majestic space* (i.e., Mountain, Desert, Woodlands, Body of Water, Cave), (2) *Haunted space* (i.e., eerie House, Cemetery, Woods, Swamp, Lonely Road), (3) *Sacred space* (i.e., Meso-America Ruin, Medieval Church, Earth-Power Spot, Ancient Greek Temple, Mosque Library), (4) *Historical space* (i.e., Ancient Rome, Great Wall of China, Stonehenge, Royal Palace, Easter Island), and (5) *Happy space* (buildings or other common structures exhibiting human face pareidolia, cf. Wang et al., 2022). In contrast, the Disenchanted images involved (1) *Outdoor space* (i.e., familiar landscapes, including fields or creeks), (2) *Residential space* (i.e., normal housing neighborhoods), (3) *Industrial space* (i.e., everyday warehouses or parking lots), (4) *Community space* (i.e., common areas in towns and cities), and (5) *Commercial space* (i.e., outside facades of malls or office buildings). For discussions about how such image contents relate to enchantment, Jungian archetypes, and general aesthetic experience, see the works of Houran and colleagues (Houran et al., 2026; Houran & Laythe, 2026).

Technical validation of the IPT also included cross-checking the accurate collection of participants' informed consent, demographic information, and psychometric data on "bound-

ary-thinness" and "encounter proneness," the latter of which feature in "Part 3: Viability of a Dual-Mode Approach to Psi Testing." Of course, researchers can substitute or add other standardized measures in transpersonal psychology or parapsychology for different research designs (see e. g., Goulding & Parker, 2001; MacDonald et al., 1995, 1999a, 1999b). For example, we plan a larger study to assess the psi performance of encounter-prone individuals – defined here as those who meet one-standard deviation (SD) thresholds on certain screening variables. Participants falling below these thresholds (or alternative scoring criteria, as needed) are routed to an auxiliary questionnaire track, whereas those who meet the inclusion criteria advance to the full IPT protocol. Individuals are given a context story and randomly assigned either to an "enchanted or disenchanted" stimuli condition, which comprised a five-trial test of putative psi using five fixed-stimulus photographs tailored to the respective condition. Between each picture selection, the participant is shown two narratives where they are allowed to select a physical direction to move in their immersive story, i. e., "ahead, behind, left, and right" choices, with the corresponding narrative following from the specific picture the participant selected. The testing protocol concludes after five psi trials (with two picture targeted narrative choices in between each trial).

2.2 Application Construction and Method

The IPT is delivered in HTML to enable efficient, web-based, and highly scalable administration. We built the program in Streamlit, hosted the code on GitHub, and configured the application to submit data to an Amazon S3 bucket, where it was saved as a CSV file for analysis. All three were password protected for extreme data protection (including individual dual verification, encryption, and passwords) regarding the data submission to the .csv residing in the Amazon S3 bucket. Because public release of the IPT's full code could inadvertently reveal some key aspects of its workings to informed study participants, its access is restricted to qualified researchers upon written request.[1]

1 "Qualified researcher" refers to an investigator with appropriate academic or professional credentials – typically someone affiliated with a university, research institute, or comparable organization – who has experience conducting human-subjects research and a legitimate scholarly interest in the IPT. Access is granted only upon written request, and recipients must agree to maintain the confidentiality of the code and protect the intellectual property it embodies, including refraining from sharing or publicly posting any sensitive implementation details.

With the assistance of ChatGPT-5 (OpenAI, 2025), we constructed a Python code that integrates a principal .py file containing initial scales and demographics, logic switches based on future application for participants meeting sufficient scoring on the accompanying psychometric measures (see below), and the remaining coding in which (1) participants are selected either for psi testing trials, based on psychometric inclusion criteria, (2) presentation (in randomized order) of either five sub-sets of five Enchanted images or five Disenchanted images (50 images total), (3) secret selection by the program using Python's "random" function which generates quasi-random functions using Python's Mersenne Twister RNG, with standard OS entropy initialization for each run of one of the five images presented as the target, across the five trials, and (4) collection of both the IPT's and participants' selections to calculate matches or psi "hits."

Further, the application leveraged additional .py files to present the AI-generated narratives based on each individual target image between each five-image psi selection trial. This process inserts two narrative-based paragraphs in which the participant is given a description of the "portal" (i.e., image) they selected and were constructed under the narrative premise that the participant had stepped "into" the image. These narrative interjections represent two exposures between each "explicit" five-picture selection psi test, creating an overall narrative and 10 total narrative choices. Each narrative presented provides participants with a choice to "move ahead, backwards, left, or right." Both files facilitated logging the participant's choice to move per one of these four directions, *and* the program's secret selection of a direction using the above-described Python "random" function to test for Implicit Psi.

AI-generated narratives associated with each image – representing eight descriptions based on each individual image totaling 400 narratives (25 enchanted images and 25 disenchanted images × 8 narratives) – were pre-specified using ChatGPT-5 and stored within a .csv for the programs to access. These were created by feeding target images to the AI program and having it first describe the image for itself using appropriate detail in written terms so that it could use its LLM to generate suitably detailed and engaging narratives. We note here the growing integration of AI both within parapsychological research (e.g., Greyson et al., 2025; Houran et al., 2025; Houran & O'Keeffe, 2025; Mossbridge et al., 2025; O'Keeffe et al., 2024) and, more broadly, in efforts to detect and counter misinformation on frontier science topics (Impey et al., 2025) – two domains that together highlight AI's expanding role in advancing methodological rigor and safeguarding knowledge.

Considering AI's current strengths and weaknesses, we purposefully elected to pre-generate the narratives for several reasons. *First*, we wished to avoid additional calls from the program to another third-party party service, thus minimizing occasional code glitches depending on the load and network availability of the AI program. *Second*, and more importantly, ChatGPT-5 (or

any similar system for that matter) shows imperfect performance, particularly with regards to repetitive or looped generation of material based on static parameters, i. e. conflicts between standard if/then looping code within the AI architecture versus the vector matrix process of the LLM.

LLMs also can occasionally exhibit "drift," meaning they can generate misaligned content based on k-temperature [i. e., a sampling scheme that first restricts next-token candidates to the top k highest-probability tokens and then applies a temperature scaling to those tokens' probabilities before sampling, allowing simultaneous control of candidate diversity (k) and randomness (temperature)] and other related variables, as well as low cosign returns on its *n*-dimensional matrix. As a result, the IPT's current iteration avoids the potential confound of AI providing non-image-oriented narrative, keeping any potential error variance due to "drifted" narrative, or failed calls to the AI controlled for. As such, the pre-generated narratives from AI, having used a consistent method and process, holds error variance constant within a known set of pre-generative narratives, immune from additional error variance being contributed due to spontaneous AI generation.

To recap, the results above reflect two types of psi tests. In the Explicit Psi test, participants view a series of five images – each either "enchanted" or "disenchanted" and rendered in high immersive quality – and select one across five trials (chance hit rate: 20%, or 1 in 5). In the Implicit Psi test, two short text narratives appear between each explicit trial, repeated five times; in each narrative, the participant chooses a direction twice from four options (chance hit rate: 25%, or 2.5 out of 10 trials). In both tests, the program secretly pre-selects a target – a picture or a direction – *before* the participant responds. This selection is never disclosed. A "hit" occurs when the participant's choice matches the program's pre-selected target; a "miss" occurs when it does not.

2.3 Checks for Randomness of Target Selection

Credible tests of putative psi require the random-selection of target stimuli and other related stimulus issues. To reiterate, all random selections for the IPT used the quasi-random Python "random" function noted earlier. Below we report the results of randomization analyses using the pilot data ($N = 126$) from "Part 3: Viability of the Dual-Mode Application for Psi Research."

2.3.1 Enchanted versus Disenchanted Conditions

As previously outlined, participants were randomly assigned either an Enchanted-stimuli or Disenchanted-stimuli condition and then directed to choose from one of five different sets of

five images associated with the particular testing condition. This produces a $p = .5$ probability of participants' assignment either to the Enchanted or Disenchanted condition.

Across the sample, 65 of 126 cases (51.6%) were assigned to the Enchanted condition and 61 (48.4%) to the Disenchanted condition. Binomial testing indicates that the observed proportion of E trials did not differ significantly from the expected chance level of .50, $p = .79$, supporting that the set assignments are consistent with a 50/50 distribution of Enchanted versus Disenchanted conditions.

2.3.2 Computer Selection for Explicit Psi

Again, once a sub-condition had been selected, five images serve as the stimuli for a five-trial test of Explicit Psi, where it would be expected that each application-selected target would be chosen with equivalent probability across each trial. Ergo, each application-selected image across a sample should be roughly equivalent to 20% (or within CI expectations) for each of the five images representing one trial across the sample.

Across the five Explicit Psi trials, we tested whether the application-selected target position (values 1–5) in each trial was consistent with a uniform random distribution (expected $p = .20$ for each position). For each of the five trials, a chi-square goodness-of-fit test compared the observed frequencies of target positions to the expected equal frequencies, with Cramér's V (reported hereafter as V) used to quantify effect size.

For Trial 1, the distribution of target positions did not differ significantly from uniform, $\chi^2(4, N = 126) = 2.02$, $p = .733$, $V = .06$ (counts: 1 = 23, 2 = 26, 3 = 22, 4 = 31, 5 = 24). The same was observed for Trial 2, $\chi^2(4, N = 126) = 1.62$, $p = .805$, $V = .06$ (1 = 26, 2 = 23, 3 = 29, 4 = 27, 5 = 21); Trial 3, $\chi^2(4, N = 126) = 6.30$, $p = .178$, $V = .11$ (1 = 20, 2 = 28, 3 = 17, 4 = 31, 5 = 30); and Trial 4, $\chi^2(4, N = 126) = 1.78$, p = .777, $V = .06$ (1 = 25, 2 = 25, 3 = 27, 4 = 20, 5 = 29). In contrast, Trial 5 showed a statistically significant deviation from the expected uniform distribution, $\chi^2(4, N = 126) = 11.38$, $p = .023$, $V = .15$, with relatively fewer "1" selections ($n = 13$) and more "2" selections ($n = 36$) than expected (other counts: 3 = 27, 4 = 22, 5 = 28).

To evaluate the overall pattern across the full IPT protocol, we pooled all five trials (Trial 1–Trial 5) into a single analysis of 630 computer selections. The combined distribution of target positions (1 = 107, 2 = 138, 3 = 122, 4 = 131, 5 = 132) did not differ significantly from a uniform random distribution, $\chi^2(4, N = 630) = 4.62$, $p = .329$, $V = .04$, indicating that, in aggregate, (noting the aberrant behavior of Trial 5) the application's target assignments were consistent with equal random selection among the five positions.

2.3.3 Computer Selection for Implicit Psi

Between each picture trial are two narratives presented with four options of which the IPT also secretly selected a "direction," i.e., when the participant is directed to move "forward, backward, left or right" at certain decision points in the immersive storyline. This represents 10 application-selected choices where each option within each narrative should also be roughly equivalent to 25% for each direction choice within the sample.

For each of the 10 narrative four option choices we examined whether the four possible target values (1–4) followed a uniform distribution (p = .25) using chi-square goodness-of-fit tests. Results indicate that most variables showed no statistically significant deviation from uniformity: Narrative1a (counts: 1 = 32, 2 = 24, 3 = 37, 4 = 33), $\chi^2(3, N = 126) = 2.83$, p = .42, V = .09; Narrative 2a (1 = 27, 2 = 27, 3 = 32, 4 = 40), $\chi^2(3, N = 126) = 3.59$, p = .31, V = .10; Narrative 2b (1 = 33, 2 = 28, 3 = 36, 4 = 29), $\chi^2(3, N = 126) = 1.30$, p = .73, V = .06; Narrative 3a (1 = 39, 2 = 34, 3 = 29, 4 = 24), $\chi^2(3, N = 126) = 3.97$, p = .27, V = .10; Narrative 3b (1 = 33, 2 = 32, 3 = 32, 4 = 29), $\chi^2(3, N = 126) = 0.29$, p = .96, V = .03; Narrative 4a (1 = 25, 2 = 34, 3 = 36, 4 = 31), $\chi^2(3, N = 126) = 2.19$, p = .53, V = .08; Narrative 4b (1 = 35, 2 = 29, 3 = 31, 4 = 31), $\chi^2(3, N = 126) = 0.60$, p = .90, V = .04; Narrative 5a (1 = 30, 2 = 33, 3 = 33, 4 = 30), $\chi^2(3, N = 126) = 0.29$, p = .96, V = .03.

In contrast, two variables showed statistically significant deviations from the expected uniform distribution: Narrative1b (1 = 44, 2 = 31, 3 = 35, 4 = 16), $\chi^2(3, N = 126) = 12.98$, p = .005, V = .19, and Narrative 5b (1 = 45, 2 = 27, 3 = 23, 4 = 31), $\chi^2(3, N = 126) = 8.73$, p = .033, V = .15, indicating moderate deviations from equal selection probabilities for those specific directional phases.

We pooled the 10 narrative variables into a single analysis (total N = 1,260 computer choices; counts: 1 = 343, 2 = 299, 3 = 324, 4 = 294) to examine the overall pattern across all directional selections. A chi-square goodness-of-fit test again compared the observed frequencies to a uniform distribution with equal expected probabilities for each of the four target values. The pooled distribution did not differ significantly from uniform, $\chi^2(3, N = 1{,}260) = 4.96$, p = .18, V = .04, indicating that in aggregate the IPT application's directional target assignments are consistent with approximately equal random selection among the four options.

2.4 Checks for Randomness of Image Presentation

Below we describe the procedures used to validate that display order of the stimulus images did not introduce systematic bias into the IPT. We first summarize the stimulus selection and norming that produced the two stimulus classes used in the task. Next, we detail the randomization algorithm and present statistical tests (contingency tables, chi-square tests of independence,

and V applied to each trial and to the pooled placements to confirm that image identity was independent of screen position or to quantify any small departures from ideal randomness.

2.4.1 Enchanted and Disenchanted Images

The IPT was designed to test our primary hypothesis that immersion and situational-enchantment enhance psi functioning, as based on prior work (Houran et al., 2024; Lange et al., 2023; Lange & Houran, 2021). Note that Houran and Laythe (2026) confirmed the differences in the Enchanted and Disenchanted themed photographs, where participants rated 50 images as "enchanted" and 50 pictures as "disenchanted" using a six-item modified version (alpha = .77, score range = 6–24) of the Visitor Experience Questionnaire (VEQ; Houran et al., 2023) to index the six environmental gestalt effects specified earlier.

We then selected from Houran and Laythe's (2026) normed collection of photographs the 25 highest-rated "enchantment" images for the Enchanted category and 25 of the lowest-rated "disenchantment" items for the Disenchanted category in our IPT protocol. Across images, those in the Enchanted condition (E; n = 25 images, total ratings N_E=387) received markedly higher ratings (M = 18.83, SD = 1.13) on environmental gestalt effects than images in the Disenchanted condition (D; n = 25 images, total ratings N_D=427; M = 14.07, SD = 1.47). An independent-samples Welch's t-test on image-level means demonstrates strong statistical significance, $t(45.10) = 12.83$, $p < .001$, 95% CI for the mean difference [4.02, 5.51], d = 3.63.

2.4.2 Psi Image Presentation Order

To avoid a selection bias, the IPT randomly shuffles the presentation order of the displayed image set. Thus, across five trials in aggregate, the displayed position order (1, 2, 3, 4, 5) of any given image position should be approximately 20%. We evaluated the effectiveness of the image-layout randomizer by treating display position and image index as two categorical variables and checked whether they were independent for each trial. For each layout variable (Trial 1 through Trial 5), we parsed the stored permutation (e.g., [1, 2, 0, 3, 4]) into five positions (Position 1–5 on the screen) and five image indices (0–4). A 5 × 5 contingency table was then constructed for each trial, where rows represented screen position and columns represented which image index appeared in that position. Under the assumption of a properly functioning randomizer, image index should be independent of position, such that each image is equally likely to appear in any position. This was tested with chi-square tests of independence for each trial ($df = (5 - 1) \times (5 - 1) = 16$), using V as the effect size. Finally, we pooled all five trials into a single 5 × 5 table (collapsing across Trials 1 through 5) and repeated the same analysis to assess the overall performance of the layout randomizer across the entire psi testing series.

Across individual trials, most layout indices were consistent with random image placement. For Trial 1, the association between screen position and image index was not statistically significant, $\chi^2(16, N = 630) = 17.86$, $p = .332$, $V = .08$, indicating no meaningful positional bias in the first trial's layouts. Similarly, Trial 3 showed no evidence of dependence between position and image index, $\chi^2(16, N = 630) = 10.48$, $p = .841$, $V = .06$, and Trial 5 was likewise nonsignificant, $\chi^2(16, N = 630) = 18.89$, $p = .274$, $V = .09$. In these three trials, the observed distributions of images across positions were consistent with the pattern expected from an unbiased randomization process.

Two trials showed statistically detectable, though small, deviations from perfect positional randomness. For Trial 2, there was a significant association between screen position and image index, $\chi^2(16, N = 630) = 28.97$, $p = .024$, $V = .11$. Likewise, Trial 4 yielded a significant chi-square, $\chi^2(16, N = 630) = 41.90$, $p < .001$, $V = .13$. These results suggest that, in Trials 2 and 4, certain images occurred in particular positions somewhat more (or less) often than would be expected under strict positional independence. However, the effect sizes were rather small ($V \approx .11$–$.13$), indicating modest deviations rather than strong or systematic ordering biases.

When all five trials were pooled into a single analysis (combining Trial 1 through Trial 5) into one 5×5 table of position by image index ($N = 3{,}150$ placements), the overall pattern showed a statistically significant but small association between screen position and image identity, $\chi^2(16, N = 3{,}150) = 36.37$, $p = .003$, $V = .05$. Thus, aggregated across all trials, the layout randomizer produced distributions that are very close to, but not perfectly indistinguishable from, what would be expected under ideal positional independence. The statistically significant result is driven primarily by subtle, small-effect deviations (notably in Trials 2 and 4), rather than a systematic lack of randomization on a practical level.

2.5 Operationalization of Implicit Psi via Narrative Decision Points

The IPT's 10 narrative decision points – each offering participants a choice among four neutral directions ("ahead, behind, left, or right") embedded directly in the second-person storyline – were deliberately constructed to operationalize Implicit Psi as low-pressure, non-conscious deviations from chance. Unlike the five explicit image-selection trials, where participants are overtly primed to "use your psychic intuition" on designated portals, the directional choices are presented solely as natural story-progression moves with no mention of testing, scoring, or paranormal performance. This covert framing leverages the immersive, gamified format to elicit automatic, intuitive selections while participants remain unaware

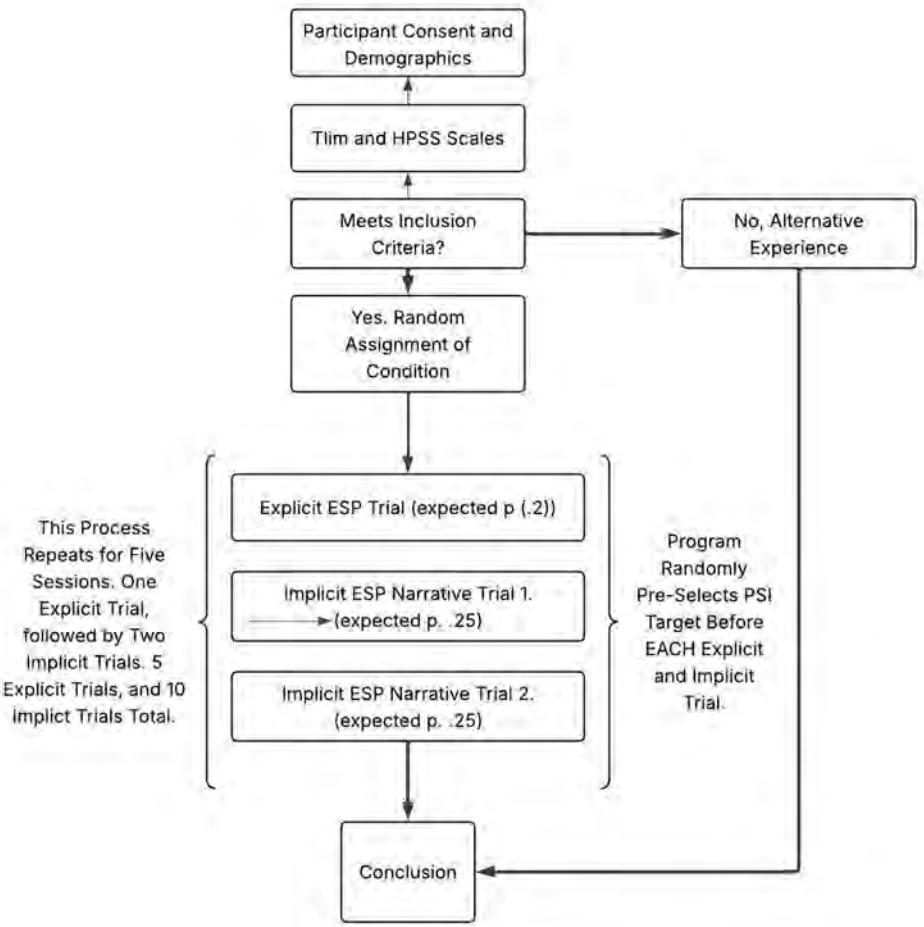
that their responses are being evaluated against a hidden random target. By design, the narratives were pre-generated with balanced, thematically equivalent descriptions across all four options (verified through internal consistency checks), eliminating systematic differences in aesthetic appeal, emotional valence, or narrative plausibility that might otherwise guide choices (see e.g., Delanoy & Solfvin, 1996).

We explicitly recognize that above-chance performance on these points could reflect non-psi mechanisms, including unconscious aesthetic preferences (e.g., favoring "forward" movement in adventurous contexts), implicit demand characteristics (e.g., participants unconsciously aligning choices with the "enchanted" framing), or subtle priming from preceding imagery and story elements. However, several safeguards reduce these alternatives: (a) true randomization of targets via the Mersenne Twister (with pooled chi-square tests confirming overall uniformity despite minor trial-level deviations), (b) absence of trial-by-trial feedback or experimenter cues, and (c) the condition-specific pattern observed only in the enchanted arm – precisely where immersive engagement was hypothesized to facilitate anomalous cognition rather than ordinary biases. These features arguably position the narrative decisions as a cleaner probe of Implicit Psi than traditional forced-choice tasks, while still allowing future preregistered replications to adjudicate residual confounds through counterbalanced options and explicit bias probes. Figure 1 provides a schematic overview of the IPT workflow, illustrating the separation between the Explicit Psi path (conscious image selections) and the Implicit Psi path (covert narrative direction choices), along with the key randomization points.

3. Pilot Study: Viability of the Dual-Mode Application for Psi Research

Pilot studies are critical in the development of new approaches by ensuring that designs or tools operate as intended, thereby supporting the practical and reliable collection or analysis of data (Westlund & Stuart, 2017). Our goal therefore was to validate that the IPT functioned as designed for actual practice and in further preparation for a large, preregistered study of "enchanted psi" to be reported separately. This latter research tests the hypothesis that putative psi performance is optimized when high-transliminality or thin-boundary individuals (i.e., "exceptional subjects") engage with stimuli or settings that are associated with numinosity or a sense of enchantment (cf. Houran et al., 2024; Lange et al., 2023; Lange & Houran, 2021). In this pilot, however, these psychometric variables were not used as exclusion criteria; instead, we collected them to confirm the functionality of the measures and to explore their general correlations with psi performance.

Figure 1
Schematic diagram of the Immersive Psi Test's (IPT) dual-mode workflow, showing the explicit psi path (conscious image selections), the implicit psi path (covert narrative direction choices), and the key randomization points.



We report all results obtained in this pilot phase for illustration purposes, but no specific hypotheses are advanced. Our pilot study's design, analysis, and research materials were not preregistered, but the protocol was reviewed and approved by the Ethics Committee at

Integrated Knowledge Systems. Moreover, we strived to follow the Journal Article Reporting Standards (Kazak, 2018) and thus describe how we determined our research samples, all data exclusions (if any), specific research questions, applicable manipulations, and all measures and data abstractions.[2]

Some empirical evidence also suggests that experimenter expectancy, belief in psi, and even unconscious processes can influence outcomes in psi research (e.g., Kennedy & Taddonio, 1976; Smith, 2003). To address this issue, we note that the IPT protocol did not require any author–participant interaction, and the authors rated their *a priori* belief in the pilot study's success as "3" on a four-point scale (4 = strong belief in the success of the experiment, 3 = moderate belief, 2 = moderate non-belief, 1 = strong non-belief) – a rating that reflected cautious optimism rather than strong conviction, driven more by concerns over the technology's stability than by doubt about Lange and Houran's (2021) basic "enchanted psi" concept.

3.1 Participants

We recruited 126 participants through the Cloud Research platform (i.e., an online participant-recruitment and data-collection platform used to obtain high-quality, pre-screened study samples) to complete the psi test component of the IPT protocol, emphasizing that this pilot effort served as an essential precursor to a larger main study of participants profiling exclusively as ostensible "exceptional subjects," i.e., those who simultaneously score "> one-*SD*" on the Revised Transliminality Scale and "> the mean" on the Haunted People Syndrome Screener measures as described below. Our convenience sample was compensated for their participation but not screened for any psychometric-based inclusion/exclusion criteria. Ages ranged broadly, with a mean age of 37.4 years (*SD* = 12.6). Gender distribution was balanced, with 48.4% identifying as female (*n* = 61), 47.6% as male (*n* = 60), and 4.0% as non-binary (*n* = 5). Participants reported a range of ethnic backgrounds. The majority identified as White/European (67.5%, *n* = 85), followed by Black/African American (11.9%, *n* = 15), Asian/Pacific (9.5%, *n* = 12), and Hispanic/Latino (7.9%, *n* = 10). A small number identified as Middle Eastern/Arab or mixed heritage (0.8%, *n* = 1) or other/multiracial categories (1.6%, *n* = 2). One participant (0.8%) did not disclose ethnicity.

2 The raw pilot data are available as a Supplementary Materials file: https://www.anomalistik.de/images/pdf/zfa/supp_mat/JAnom26-1_Laythe-Houran_Supplemental-Data.csv

3.2 Psychometric Measures

Revised Transliminality Scale (RTS: Lange et al. 2000) is a 17-item, T/F, Rasch (1960/1980) scaled measure of "hypersensitivity to psychological material originating in (a) the unconscious, and/or (b) the external environment" (Thalbourne & Maltby, 2008, p. 1618). The items specifically span magical ideation, mystical experience, absorption, hyperesthesia, manic experience, dream interpretation, and fantasy proneness, all of which share a common underlying factor. Transliminality is thus a perceptual-personality variable that incorporates Hartmann's (1991) mental boundary construct and the notion of sensory processing sensitivity (Aron et al., 2012). RTS scores (M = 25, SD = 5) – with a Rasch reliability of .82 (Lange et al., 2000), retest reliability of .88 over an average of seven weeks (Thalbourne, 2000 – consistently predict different anomalous experiences, syncretic cognitions, and lower physiological thresholds (Evans et al., 2019; Irving et al., 2024; Ortega et al., 2025; Rosen et al., 2023; Roxburgh et al., 2024; Swami et al., 2024). Transliminality also correlates positively with both paranormal belief and putative psi performance (e.g., Thalbourne & Houran, 2003; Thalbourne & Storm, 2012; Ventola et al., 2019).

Haunted People Syndrome-Screener (HPSS; Lange & Houran, 2024) consists of six items to be rated four-point Likert scales anchored by "Strongly Disagree" (scored 0) and "Strongly Agree" (scored 3). These cover the presence of recurrent anomalous experiences and four of the five recognition patterns of an "encounter-prone" psychometric profile (Laythe et al., 2021, 2022), i.e. (1) Thin Boundary Functioning, (2) Dis-ease States, (3) Perceptual Contagion (i.e., covering both event flurries and diverse perceptions), and (4) Sense-Making Attributions (i.e., narrative reality based on general ideological beliefs). The Rasch-scaled scores (reliability = .80) range from 37.1 to 71.2, with a mean of 50 and standard deviation = 10. HPSS scores also strongly and positively predict scores on the Survey of Strange Experiences (attenuation corrected correlation = 0.78, $p < .001$). Notably, an encounter-prone profile maps directly onto participant characteristics consistently associated with elevated psi performance across multiple experimental paradigms, insofar as Baptista et al. (2015) found that individuals with prior anomalous experience, psi-conducive beliefs, and mental discipline practice produced significantly higher hit rates than unselected participants in ganzfeld studies, and that analogous selection effects were replicated in forced-choice ESP databases, underscoring the value of HPSS scores as an theoretically grounded inclusion criterion for psi research.

3.3 Procedure

We refer readers to the previous section for specific details of the survey and procedure. Here we reiterate that after informed consent and demographics collection, participants completed the

HPSS and transliminality measures but were automatically routed to the psi testing part of the protocol. Participants were randomly assigned to either an Enchanted-Stimuli or Disenchanted-Stimuli testing condition. The Enchanted-Stimuli Condition instructs participants to read detailed descriptions of five types of "enchanted" geographies (i.e., Majestic, Haunted, Sacred, Historical, or Happiness) and then select "*Which location sounds to you as the most intriguing to visit*." The IPT then populates the images that correspond to the participant's selection. The Disenchanted-Stimuli Condition similarly directs participants to read detailed descriptions of five types of "everyday" geographies (i.e., Basic Garden, New House, Warehouse, Mile Marker, or Worn Building Exterior) and then indicate "*Which location sounds to you as the most bor*ing *to visit*." The IPT then populates the corresponding images to match the "disenchanted" selection. This process aims to tailor the testing stimuli to each participant's expectations about the enchanting or disenchanting qualities of specific environments.

During the IPT, a participant engages with a storyline from ChatGPT-5 (OpenAI, 2025) that has the user choose one out of five stimulus pictures within their preferred enchanted category (e.g., the Haunted or Majestic setting) or the disenchanted category they deemed most mundane (e.g., the Residential or Industrial setting), with one being randomly and covertly selected by the AI program in real-time as the "psi target." This selection process repeats five times during the evolving storyline, thus yielding a five-trial test of putative psi with an expected score of 1/5 or 20%. We designed this efficient and scalable AI-facilitated narrative format to mimic the perceptual or experiential features of immersive-type experiences involving "uncanny" contents or contexts (Annett et al., 2016; Hill et al., 2018; Houran et al., 2023) but without the logistical or methodological drawbacks of fieldwork studies or free-response designs.

Between each image selection psi task, the IPT produced two narratives with a four choice option of "direction" ("ahead, behind, left, or right") which the computer also randomly and covertly selected, to create either an immersive storyline for (1) an Enchanted-Stimuli Condition, where positive descriptors are used along with a set of emotionally-laden settings (i.e., Majestic Landscape, Haunted House, Sacred Space, Historical Landmark, or Happy-Face Buildings via pareidolia) versus (2) the Disenchanted-Stimuli Condition involving mundane descriptors for commonplace settings (i.e., Basic Garden, New House, Warehouse, Mile Marker, or Worn Building Exterior). This implicit psi test carries a random expectation of 2.5/10 or 25%.

3.4 Results

The analyses below address three issues central to protocol validation: (1) whether the randomization and stimulus-assignment procedures functioned as intended; (2) whether partic-

ipants in the enchanted and disenchanted conditions performed differently; and (3) whether either condition produced psi scores reliably above chance. Each subsection pairs the primary statistical result with a plain-language interpretation to aid accessibility. Note that all analyses and interpretations concerning Implicit Psi are strictly exploratory and are intended solely for hypothesis generation rather than for claiming reliable detection of anomalous effects.

3.4.1 Preliminaries

Our analysis only examines hit rate on the IPT as a function of Enchanted vs. Disenchanted images and reinforcing narratives, given that we did not screen participants for higher scores on the HPSS and RTS. Given previous work with enchantment, transliminality, and putative psi (Houran et al., 2024; Lange et al., 2023; Lange & Houran, 2021), we would expect elevated levels of both transliminality and encounter-proneness in conjunction with an enchanted environment to produce psi scores beyond chance-levels. We again do not specify hypotheses, noting that any significant effects would imply that situational-enchantment alone is (in part) an independent variable (regardless of one's inherent traits or abilities) which modulates psi scoring. Finally, understand that we are applying a broad analysis across all Enchanted or Disenchanted conditions, leaving potential differences in each of the five subsets for future scrutiny and a larger sample.

3.4.2 Explicit vs Implicit Psi Scores Relative to Enchanted and Disenchanted Images

Analysis revealed that the summed psi scores were not normally distributed. Specifically, Shapiro-Wilk tests showed that Explicit Psi scores deviated substantially from normality in the full sample, $W(126) = 0.83$, $p < .001$, and Implicit Psi likewise departed from normality, $W(126) = 0.93$, $p < .001$. Parallel Shapiro-Wilk tests within the Enchanted (E) and Disenchanted (D) conditions also indicated non-normality for both Explicit and Implicit Psi scores (all p's $\leq .003$). Given the former, subsequent analysis employed Mann-Whitney U tests were used as the primary between-groups analyses, with U, standardized z, exact two-tailed p values, and effect sizes reported as r.

For Explicit (conscious) Psi performance, participants in the Enchanted condition had a mean hit rate of $M = 0.89$, $SD = 0.81$ (med. = 1, IQR = 0–1, $n = 65$), whereas those in the Disenchanted condition had a hit rate of $M = 0.97$, $SD = 0.80$ (med. = 1, IQR = 0–1, $n = 61$). A Mann-Whitney U test indicated that Explicit Psi scores did not differ significantly between the E and D conditions, $U = 1873.00$, $z = -0.57$, $p = .57$, $r = .05$. Thus, there was no evidence that the global Enchanted versus Disenchanted manipulation influenced Explicit Psi performance in our pilot sample with an open distribution of HPSS and RTS scores.

For Implicit (non-conscious) Psi performance, the participants in the Enchanted condition had a mean hit rate of M = 2.97, SD = 1.77 (med. = 3, IQR = 2–4, n = 65), whereas participants in the Disenchanted condition had a hit rate of M = 2.43, SD = 1.43 (med. = 2, IQR = 1–3, n = 61). Although the Enchanted group showed somewhat higher Implicit Psi scores descriptively, a Mann–Whitney U test indicated that this difference was not statistically significant, U = 1660.50, z = -1.61, p = .11, r = .14. These results indicate a small but nonsignificant tendency toward increased Implicit Psi performance in the Enchanted condition.

3.4.3 Aggregated Psi Scores Relative to Chance Expectations

Because psi performance was recorded as "counts of hits" over a fixed number of trials (five explicit trials and 10 implicit trials per participant), we treated each trial as a Bernoulli event (hit vs. miss) and used exact binomial tests to compare observed hit rates against their theoretical chance levels. For Explicit Psi, chance was defined as p_0 = .20 (one hit in five trials on average). For Implicit Psi, chance was defined as p_0 = .25 (2.5 hits in 10 trials on average). Within each condition (Enchanted vs Disenchanted) and in the total sample, we summed hits across participants and divided by the total number of trials to obtain an overall hit rate $\hat{p}$. We then conducted one-sided exact binomial tests with the directional alternative H_1: $p > p_0$ ("above-chance"), and we report effect sizes as Cohen's h.

For Explicit Psi, in the Enchanted (E) condition, participants produced 58 explicit hits across 325 trials, yielding an overall hit rate of $\hat{p}$ = .18(95% *CI* [.14, .22]). A one-sided exact binomial test indicated that this rate did not exceed chance, exact p = .85, Cohen's h = -0.06, suggesting psi performance was slightly below but statistically indistinguishable from the expected .20 hit probability. In the Disenchanted (D) condition, participants had 59 hits across 305 trials $\hat{p}$ =.19, 95% *CI* [.15, .24], with no evidence of above-chance performance, exact p = .64, h = -0.02. Collapsing across conditions, the total sample showed 117 explicit hits out of 630 trials, $\hat{p}$ =.19, 95% *CI* [.16, .22], again consistent with chance, exact p = .83, h = -0.04. Thus, Explicit Psi performance was not significantly above-chance in either condition or in the combined sample.

For Implicit Psi, E participants produced 187 hits across 650 trials, corresponding to an overall hit rate of $\hat{p}$ = .29, 95% *CI* [.25, .32]. This rate was significantly above the chance expectation of .25, exact one-tailed binomial p = .016, Cohen's h = 0.09, indicating a very small but reliable elevation in Implicit Psi performance within the Enchanted condition. This magnitude nonetheless falls below conventional small-effect benchmarks (h = 0.20). For comparison, this is in the range of effect sizes commonly observed in implicit cognition paradigms and subliminal priming studies (e.g., $h \approx 0.05$–0.15; Greenwald et al., 2003), as well as prior forced-choice psi meta-analyses (Honorton & Ferrari, 1989; Storm et al., 2012). The result is therefore consistent

with the modest, statistically fragile effects typical of anomalous cognition research and should not be over-interpreted on the basis of this single pilot sample.

In contrast, the D participants had 148 hits across 610 trials, $\hat{p}$ = .24, 95% *CI* [.21, .28], which did not differ from chance, exact p = .68, h = -0.02. When all participants were combined, there were 335 implicit hits out of 1,260 trials, $\hat{p}$ = .27, 95% *CI* [.24, .29], yielding a nonsignificant trend for above-chance performance, exact p = .10, h = 0.04. Overall, these binomial tests suggest that Implicit Psi scores were significantly above-chance only within the Enchanted condition, whereas the Disenchanted condition and the full sample did not show statistically reliable above-chance implicit performance.

3.4.4 Boundary-Thinness, Encounter-Proneness, and Response Time Relative to Explicit Psi Scores

Spearman rank-order correlations were computed among scores on the HPSS (encounter-proneness), RTS (boundary-thinness) Explicit Psi, Implicit Psi, and average Explicit Psi Response Time. Because these variables had non-normal distributions, Spearman's ρ was used as a nonparametric measure of association. All correlations reported here are unattenuated and based on N = 126 (df = 124).

There was a moderately strong and positive association between HPSS and RTS scores [$\rho(124) = .52$, $p < .001$], which replicates Laythe et al.'s (2018) core finding that boundary-thinness correlates with haunt-type or encounter experiences. In contrast, HPSS scores were not related to Explicit Psi ($\rho = -.14$, $p = .11$), Implicit Psi ($\rho = .02$, $p = .78$), or Explicit Psi Response Time ($\rho = .10$, $p = .27$). Transliminality, contrary to other evidence (Ventola et al., 2019, pp. 157–160) likewise showed no significant correlations with Explicit Psi ($\rho = .06$, $p = .51$), Implicit Psi ($\rho = -.01$, $p = .89$;), or Explicit Psi Response Time ($\rho = .15$, $p = .10$). Explicit and Implicit Psi scores also were unrelated, $\rho(124) = -.01$, $p = .88$, and neither index of putative psi showed a meaningful association with Explicit Psi time (Explicit: $\rho = .03$, $p = .74$; Implicit: $\rho = -.01$, $p = .92$). Overall, aside from the expected strong linkage between transliminality and HPSS scores, psi performance and response latency were not detectably correlated with trait scores or with each other in this pilot sample.

4. Conclusions, Caveats, & Future Directions

4.1 Conclusions

The IPT performed as designed. Randomization checks were broadly satisfactory, stimulus valence separation was large and reliable, and the protocol engaged participants within its intended time frame. Pilot psi results were mixed: Explicit Psi performance was at chance across both conditions, whereas Implicit Psi in the E condition showed a small above-chance deviation.

Aggregate trends for the whole sample leaned toward significance ($p \approx .10$), driven primarily by the performance in the E group. Ancillary analyses provided no evidence that putative psi was significantly influenced by the psychometric variables measured here. Scores on Encounter-proneness (via the HPSS) and boundary-thinness (via the RTS) were positively correlated as expected (r =.52), but neither HPSS, RTS, nor Elapsed Time showed reliable associations with IPT scores.

Because the pilot intentionally sampled typical-range scores rather than "exceptional subjects" (i.e., high scorers on RTS or HPSS), our pilot data do not robustly adjudicate hypotheses about moderation by trait susceptibility to, or ability for, psi-type phenomena (cf. Rock et al., 2023, Table 1, p. 14). These collective outcomes are therefore best understood as proof-of-concept findings from a convenience sample not screened for the trait profile targeted in the planned main study; they should not be interpreted as confirmatory evidence of psi.

4.2 Caveats and Considerations

Interpretively, the disconnect between null Explicit Psi (image-based) performance and positive Implicit Psi (narrative-based) performance suggests a plausible mechanism in which the E-related pictorial context frames participant experience while explicit selection pressure suppresses performance. In this account, covert narrative choices – made without participants' awareness that they are being evaluated for psi – allow low-pressure, intuitive responses to manifest more clearly. This pattern aligns with prior parapsychological findings: relaxed, passive attention states (rather than effortful concentration) are associated with higher psi performance in ganzfeld experiments (Honorton, 1977; Honorton & Harper, 1974), with theoretical discussions of an "effort paradox" in anomalous cognition (Braud, 2003), and with ethnographic observations that ritual magic practitioners often view excessive conscious striving as counterproductive (Luhrmann, 1989).

However, this interpretation must be tempered by an important caveat. Statistically significant above-chance performance in a covert decision task is consistent with – but does not confirm – anomalous cognition. Similar patterns could arise from unconscious response biases (e.g., preferences for particular narrative directions), demand characteristics, differential attentional or working-memory demands across task formats, or systematic non-randomness in target assignment. The present design cannot adjudicate among these possibilities. Accordingly, we use the term "implicit psi" as a descriptive label for above-chance performance on the covert measure, not as a claim that the effect is free from conventional explanations. Disentangling genuine anomalous cognition from these alternatives will require preregistered, double-blind protocols with explicit bias-detection checks and orthogonal control conditions.

Beyond these interpretive considerations, several methodological limitations also warrant attention. A small number of individual trials showed minor departures from perfect randomness; even slight systematic patterns could influence sensitive significance tests and therefore require close monitoring in future confirmatory work. Presentation-order tests reached significance in the large aggregate sample despite a negligible effect size – likely a consequence of sample size – so this should be treated as a minor technical flag rather than evidence of substantive bias, though it still motivates tighter randomization controls. The pilot sample also did not specifically recruit high scorers on the RTS or HPSS, limiting strong inferences about trait moderation. Finally, the exploratory and underpowered nature of several comparisons means that observed effect sizes, particularly for Implicit Psi, should be considered provisional and in need of preregistered replication. Consequently, the psi-related analyses are best interpreted as proofs-of-concept demonstrating the feasibility of collecting larger samples for high-powered, confirmatory designs.

4.3 Future Directions

The IPT advances psi methodology by providing a scalable, open-science-aligned framework that merges forced-choice precision with narrative immersion. If the Implicit Psi signal observed here replicates under preregistered conditions with samples recruited for high boundary-thinness and encounter-proneness, the findings would support a context-dependent dissociation between Explicit and Implicit Psi responding – a pattern with implications for models of attention, anomalous cognition, and consciousness. Null replication, equally, would clarify the boundaries of the protocol and rule out several methodological confounds. Either of these outcomes advances cumulative knowledge. Future work should (1) preregister primary and secondary hypotheses; (2) employ stratified sampling targeting high transliminality and

encounter-proneness; (3) strengthen randomization with block-randomization diagnostics run prior to data collection; (4) include within-subject cross-over designs to directly compare explicit and implicit response formats while controlling for demand characteristics and arousal; and (5) include sensitivity analyses (e.g., variance-subtraction, robustness checks) to quantify how minor non-randomness possibly affects primary inferences.

5. General Discussion

We support diverse experimental tools and designs over a single methodological standard (cf. Wertz, 2025); therefore, our first-generation application is intended to complement – not replace – other testing approaches for putative psi. As Morse (2003) observed, convergent validity improves when mixed or multiple methods are used because findings from varied approaches tend to be more robust than those based on a single technique This principle is especially relevant in experimental psi research where the elusive, situational, and person-sensitive nature of the phenomena argues for methodological pluralism. We implemented that philosophy in our IPT application, which appears to work as designed: the protocol proved operationally feasible, engaged participants, produced interpretable data, and functioned as an effective platform for both exploratory observation and preregistered follow-up testing. Importantly, the IPT illustrates a methodological innovation in a dual-mode design that merges the objectivity of forced-choice scoring with the ecological richness of narrative immersion; this hybrid approach thus parallels earlier advances such as the ganzfeld procedure but extends them into scalable digital environments aligned with open-science principles. By combining transparent scoring rules, automated randomization, and preregistered analytic potential with genuine participant engagement and affective salience, the IPT directly addresses several long-standing criticisms and limitations of traditional parapsychological protocols.

Mayer (2024) and Houran (2024) have similarly endorsed methodological flexibility in psi research. Mayer proposed a multi model, multi method approach in which contrasting theoretical frameworks (e.g., quantum informed vs. metaphysical accounts) prescribe different experimental tactics and a shared reliance on naturalistic mixed measures (e.g., behavioral logs and random number generators) tied to clearer theory to design mapping. Houran framed the trade off as tightly controlled "Trojan Horse" experiments versus ecologically valid "Trojan Mice" field studies, defending lower control, real world tests as productive sources of hypotheses for rigorous follow up. Together these perspectives trace parapsychological research from classical laboratory work to multi sensor field initiatives and point to a practical workflow: use open, multimodal observation to surface signals and modulators, then subject promising patterns to preregistered,

sensor rich experiments that isolate mechanisms and boundary conditions while employing contemporary safeguards such as registered reports and improved statistical practice, as well as remain attentive to experimenter and expectancy effects (cf. Wiseman & Schlitz, 1997). Reciprocal flexibility is arguably the most efficient path to scientific clarity on psi related phenomena.

We would expect that both (1) marked shifts in scores and (2) beyond chance level scores associated with our IPT's dual mode approach will correspond to an interactionist effect, i. e., positive results are more consistent or stronger when thin boundary individuals engage with enchanted stimuli. This prediction follows from prior findings consistent with the theorized success formula of "Transliminality × Paranormal Belief × Situational Enchantment" (Houran et al., 2024; Lange et al., 2023; Lange & Houran, 2021), which challenges the prevailing assumption that factors such as *paranormal belief* (the classic "sheep-goat effect"; Storm & Tressoldi, 2017), *thin mental boundaries* (Thalbourne & Storm, 2012), or *numinous target stimuli* (Krippner et al., 2019) are, on their own, sufficient to consistently produce positive psi outcomes. This interactionist prediction is further supported by Baptista et al.'s (2015) meta-analytic evidence that participant selection – particularly combinations of prior anomalous experience, psi-conducive belief, and mental discipline practice – is the dominant moderator of effect size across multiple psi paradigms, and is directly anticipated by Parker et al.'s (1998) ganzfeld findings, in which psi-hitting was jointly predicted by paranormal experience, sheep-goat belief, magical ideation, feeling-type personality, and emotional target characteristics, with no single factor sufficient on its own. Reliable psi outcomes, it seems, emerge at the intersection of the right person, the right mindset, and the right experimental context. The IPT framework thus aims to enhance experimental rigor in the study of potentially low probability, high noise phenomena with its automated instrumentation, standardized blinding procedures, and robust statistical controls to minimize bias and maximize reproducibility.

Beyond parapsychology, the IPT contributes to wider debates on cognition and consciousness. The putative Implicit Psi effects observed here resonate with research on unconscious information processing, attentional modulation, and narrative engagement, suggesting that immersive testing may serve as a novel platform for probing subtle cognitive biases and anomalous experiences. Translationally, similar immersive designs could be adapted to study altered states, ecological decision making, or implicit bias, thereby further situating psi research within mainstream psychological inquiry. Future work should certainly prioritize preregistered, multi lab replications with stratified sampling of high-transliminal participants, tighter randomization controls, and integration of psychophysiological measures – such as electroencephalogram (EEG), heart rate variability (HRV), or functional magnetic resonance imaging (fMRI) – to probe underlying mechanisms (e. g., Gosseries et al., 2024). Multiple data pipelines and collab-

orative infrastructures will be essential to determine whether implicit deviations reflect genuine anomalies or procedural artifacts. Regardless of outcome, our IPT offers a replicable, transparent, and engaging framework that can help to bridge methodological rigor with a reasonable degree of ecological validity. It is therefore a pragmatic, alternative approach for researching potential mediators, moderators, or outcomes of immersive-type experiences (psi oriented or not) without the many hurdles associated with field-based designs.

If psi phenomena ultimately prove robust, replicable, and causally interpretable, the implications for theories of mind would be profound (for discussions on these criteria, see e.g., Houran et al., 2017, 2018). The dominant view of consciousness as strictly brain bound would be seriously challenged, demanding renewed attention to alternative frameworks – such as neutral monism, panpsychism, and information centric models – that accommodate so-called "nonlocal or extended" aspects of sensation and perception (Wahbeh et al., 2022). Such a paradigm shift would necessitate revisions to cognitive and neuroscientific theory, stimulate the development of new experimental ontologies that integrate subjective experience with objective measurement, and catalyze interdisciplinary research spanning philosophy, neuroscience, physics, and information science (Vernon, 2020). Beyond these conceptual ramifications, empirical validation of psi would transform how perception, agency, and the boundaries-of-self are understood, opening pathways to novel technologies and interventions grounded in expanded models of mind (Williams, 2013). Such evidence would oblige science to broaden its explanatory framework – conceiving consciousness as potentially extending beyond individual brains – and, in doing so, to reconsider the ethical, technological, and existential dimensions of what it means to be human (Friedman et al., 2021).

Taken altogether, we propose that the IPT is not merely a tool for parapsychological research but a working model for a new class of experimental instrumentation capable of probing questions that traditional paradigms cannot always access. By fusing methodological rigor with the capacity for immersive engagement, the present AI-supported technology provides a scalable platform that can both generate exploratory signals and support preregistered confirmatory tests. Whether or not future replications confirm the small Implicit Psi effects observed here, the IPT advances the field by offering a transparent, adaptable framework for cumulative science and by demonstrating how innovative methodologies can expand the empirical reach of consciousness research itself. In this sense, the IPT can help to reframe research on putative psi not at the fringe, but the frontier (cf. Houran & Bauer, 2022) – inviting the wider scientific community to pursue, with an appropriate balance of caution and openness, the kinds of experimental challenges that will determine the mechanisms and limits of human cognition and the broader mind.

Acknowledgements

This work was part of a larger study of "enchanted psi" supported by the BIAL Foundation (bursary #05/24). We thank the peer reviewers for helpful comments on an earlier draft.

References

Annett, M., Lakier, M., Li, F., Wigdor, D., Grossman, T., & Fitzmaurice, G. (2016). The living room: Exploring the haunted and paranormal to transform design and interaction. In M. Froth (General Chair), *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (DIS '16) (pp. 1328–1340). Association for Computing Machinery. https://doi.org/10.1145/2901790.2901819

Aron, E. N., Aron, A., & Jagiellowicz, J. (2012). Sensory processing sensitivity: A review in the light of the evolution of biological responsivity. *Personality and Social Psychology Review, 16*, 262–282. https://doi.org/10.1177/1088868311434213

Baker, I. S., & O'Keeffe, C. (2007). Ethical guidelines for the investigation of haunting experiences. *Journal of the Society for Psychical Research, 71*(4), 216–229.

Bandrowski, A., Martone, M. E., & Poline, J.-B. (2023). Promoting reusable and open methods and protocols. *PLoS Biology, 21*, Article e3002835. https://doi.org/10.1371/journal.pbio.3002835

Baptista, J., Derakhshani, M., & Tressoldi, P. E. (2015). Explicit anomalous cognition: A review of the best evidence in ganzfeld, forced choice, remote viewing and dream studies. In E. Cardeña, J. Palmer, & D. Marcusson-Clavertz (Eds.), *Parapsychology: A handbook for the 21st century* (pp. 192–214). McFarland & Co.

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology, 100*, 407–425. https://doi.org/10.1037/a0021524

Bem, D. J., & Honorton, C. (1994). Does psi exist? Replicable evidence for an anomalous process of information transfer. *Psychological Bulletin, 115*, 4–18. https://doi.org/10.1037/0033-2909.115.1.4

Bem, D., Tressoldi, P., Rabeyron, T., & Duggan, M. (2015). Feeling the future: A meta-analysis of 90 experiments on the anomalous anticipation of random future events. *F1000Research, 4*, Article 1188. https://doi.org/10.12688/f1000research.7177.2

Bendall, R. C. A., Royle, S., Dodds, J., Watmough, H., Gillman J. C., Beevers D., Cassidy S., Short, B., Metcalfe, P., Lomas, M. J., Graham-Kevan, D., & Gregory, S. E. A. (2025). The Salford Nature Environments Database (SNED): An open-access database of standardized high-quality pictures from natural environments. *Behavior Research Methods, 57*, Article 21. https://doi.org/10.3758/s13428-024-02556-4

Bierman, D. J., Spottiswoode, J. P., & Bijl, A. (2016). Testing for questionable research practices in a meta-analysis: An example from experimental parapsychology. *PLoS One, 11*, Article e0153049. https://doi.org/10.1371/journal.pone.0153049

Braud, W. G. (2003). *Distant mental influence: Its contributions to science, healing, and human interactions*. Hampton Roads Publishing.

Buckley, P. & Doyle, E. (2016). Gamification and student motivation. *Interactive Learning Environments, 24*, 1162–1175. https://doi.org/10.1080/10494820.2014.964263

Cardeña, E. (2018). The experimental evidence for parapsychological phenomena: A review. *American Psychologist, 73*, 663–677. https://doi.org/10.1037/amp0000236

Delanoy, D. L., & Solfvin, J. F. (1996). Exploring psychological variables of free-response ESP targets and their relationships to psi-scoring. In E. May (Ed.), *Proceedings of the Parapsychological Association 39th Annual Convention, Supplement* (pp. 1–15). Parapsychological Association.

Drinkwater, K., Massullo, B., Dagnall, N., Laythe, B., Boone, J., & Houran, J. (2022). Understanding consumer enchantment via paranormal tourism: Part I–Conceptual review. *Cornell Hospitality Quarterly, 63*, 195–215. https://doi.org/10.1177/1938965520967920

Escolà-Gascón, A., Wright, A., & Houran, J. (2022). 'Feeling' or 'sensing' the future again? Testing anomalous cognitions in clinical versus healthy populations. *Heliyon, 8*, Article e11303. https://doi.org/10.1016/j.heliyon.2022.e11303

Evans, J., Lange, R., Houran, J., & Lynn, S. J. (2019). Further psychometric exploration of the transliminality construct. *Psychology of Consciousness: Theory, Research, and Practice, 6*, 417–438. https://doi.org/10.1037/cns0000163

Friedman, H. L., Radin, D., & Krippner, S. (2021). Parapsychology and transpersonal psychology in dialogue: Could these two movements be brought into better alignment? *International Journal of Transpersonal Studies, 40*, 123–135. https://doi.org/https://doi.org/10.24972/ijts.2021.40.1.123

Galak, J., LeBoeuf, R. A., Nelson, L. D., & Simmons, J. P. (2012). Correcting the past: Failures to replicate psi. *Journal of Personality and Social Psychology, 103*, 933–948. https://doi.org/10.1037/a0029709

Gosseries, O., Marie, N., Lafon, Y., Bicego, A., Grégoire, C., Oswald, V., & Vanhaudenhuyse, A. (2024). Exploration of trance states: Phenomenology, brain correlates, and clinical applications. *Current Opinion in Behavioral Sciences, 58*, Article 101400. https://doi.org/10.1016/j.cobeha.2024.101400

Goulding, A., & Parker, A. (2001). Finding psi in the paranormal: Psychometric measures used in research in paranormal beliefs/experiences and in research on psi-ability. *European Journal of Parapsychology, 16*, 73–101.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 8*, 197–216. https://doi.org/10.1037/0022-3514.85.2.197

Greyson, B., Long, J., Holden, J. M., Jourdan, J. P., King, R. A., Mays, S., Mays, R., Rivas, T., Tassell-Matamua, N., van Lommel, P., Woollacott, M., & Tressoldi, P. (2025). The Veridical Near-Death Experience Scale: Construction and a first validation with human and artificial raters. *Frontiers in Psychology, 16*, Article 1661390. https://doi.org/10.3389/fpsyg.2025.1661390

Hartmann, E. (1991). *Boundaries in the mind: A new psychology of personality*. Basic Books.

Hill, S.A., O'Keeffe, C., Laythe, B., Dagnall, N., Drinkwater, K., Ventola, A., & Houran, J. (2018). "Meme-spirited": I. A VAPUS model for understanding the prevalence and potency of ghost narratives. *Australian Journal of Parapsychology, 18*(2), 117–152.

Honorton, C. (1977). Psi and internal attention states. In B.B. Wolman (Ed.), *Handbook of parapsychology* (pp. 435–472). Van Nostrand Reinhold.

Honorton, C., Berger, R.E., Varvoglis, M.P., Quant, M., Derr, P., Schechter, E.I., & Ferrari, D.C. (1990). Psi communication in the ganzfeld: Experiments with an automated testing system and a comparison with a meta analysis of earlier studies. *Journal of Parapsychology, 54*, 99–139.

Honorton, C., & Ferrari, D.C. (1989). "Future telling:" A meta analysis of forced choice precognition experiments, 1935–1987. *Journal of Parapsychology, 53*, 281–308.

Honorton, C., & Harper, S. (1974). Psi-mediated imagery and ideation in an experimental procedure for regulating perceptual input. *Journal of the American Society for Psychical Research, 68,* 156–168.

Houran, J. (2024). Frontier science with 'dirty test tubes' – flaw or feature? (Introduction to the special subsection). *Journal of Scientific Exploration, 38*, 297–298. https://doi.org/10.31275/20243425

Houran, J., & Bauer, H.H. (2022). "Fringe science" – A tautology, not pariah [editorial]. *Journal of Scientific Exploration, 36*, 207–217. https://doi.org/10.31275/20222527

Houran, J., Katz, D.L., Williamson, J., Koren, S.A., Wahbeh, H., & Woollacott, M. (2025). Magic flights or mind's eye? Further quali-quantitative explorations of dimensional-slip narratives. *Journal of Anomalistics, 25*, 61–118. https://doi.org/10.23793/ZFA.2025.061

Houran, J., Lange, R., & Hooper, D. (2018). Cross-examining the case for precognition: Commentary on Mossbridge and Radin (2018). *Psychology of Consciousness: Theory, Research and Practice, 5*, 98–109. https://doi.org/10.1037/cns0000126

Houran, J., & Laythe, B. (2026). Still images, shifting minds: Testing the transformative effects of evocative visuals. *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication. https://doi.org/10.1037/aca0000860

Houran, J., Laythe, B., & Houran, B.M. (2026). Still images as portals to immersive and transformative experience: A narrative review. *Journal of Transpersonal Psychology, 57*, 158–187.

Houran, J., Laythe, B., Lange, R., Hanks, M., & Ironside, R. (2023). Immersive study of Gestalt variables in uncanny geographies. *Journal of the Society for Psychical Research, 87*, 65–100.

Houran, J., Laythe, B., & Little, C. (2024). Confirmatory study of anomalous experiences in enchanted spaces. *Journal of Anomalistics, 24*, 396–426. https://doi.org/10.23793/ZFA.2024.396

Houran, J., Lynn, S.J., & Lange, R. (2017). Commentary on Stokes' (2017) quest for "white crows" in spontaneous cases of psi. *Australian Journal of Parapsychology, 17*, 61–88.

Houran, J., Maraldi, E.O., & Houran, B.M. (2025). Debating member-checks of exceptional human experiences. *International Journal of Transpersonal Studies, 44*, 10–28. https://doi.org/10.24972/ijts.2025.44.1.10

Houran, J., & O'Keeffe, C. (2025). Prototyping haunted people syndrome: Artificial intelligence-generated narrative, structured interview, and clinical validation. *Psychology of Consciousness: Theory, Research, and Practice*. Advance online publication. https://doi.org/10.1037/cns0000447

Hyman, R. (1985). The ganzfeld psi experiment: A critical appraisal. *Journal of Parapsychology, 49*, 3–49.

Hyman, R., & Honorton, C. (1986). A joint communiqué: The psi ganzfeld controversy. J*ournal of Parapsychology, 50*, 351–364.

Impey, C., Danehy, A., Wenger, M., & Buxner, S. (2025). Using machine learning to detect scientific misinformation. *Astronomy Education Journal, 5*, Article 138aer. https://doi.org/10.32374/AEJ.2025.5.1.138aer

Irving, A. J., Nikolova, N., Robinson, S., Ionita, I., Kelly, S. W., Kirsch, I., Mazzoni, G., Venneri, A., & McGeown, W. J. (2024). The relationship between transliminality, hypnotic and imaginative suggestibility, and other personality traits. *Acta Psychologica, 243*, Article 104125. https://doi.org/10.1016/j.actpsy.2024.104125

Jaramillo-Mediavilla, L., Basantes-Andrade, A., Cabezas-González, M., & Casillas-Martín, S. (2024). Impact of gamification on motivation and academic performance: A systematic review. *Education Sciences, 14*, Article 639. https://doi.org/10.3390/educsci14060639

Katz, D. L., Grgić, I., Tressoldi, P., & Fendley, T. W. (2021). Effects of background context for objects in photographic targets on remote viewing performance. *Journal of Scientific Exploration, 34*, 752–787. https://doi.org/10.31275/20212273

Kazak, A. E. (2018). Journal article reporting standards [editorial]. *American Psychologist, 73*, 1–2. https://doi.org/10.1037/amp0000263

Kekecs, Z., Palfi, B., Szaszi, B., Szecsi, P., Zrubka, M., & Aczel, B. (2023). Raising the value of research studies in psychological science by increasing the credibility of research reports: The transparent psi project. *Royal Society Open Science, 10*, Article 191375. https://doi.org/10.1098/rsos.191375

Kennedy, J. E. (2016). Is the methodological revolution in parapsychology over or just beginning? *Journal of Parapsychology, 80*, 156–168.

Kennedy, J. E., & Taddonio, J. L. (1976). Experimenter effects in parapsychology: The importance of the experimenter's expectations and attitudes. *Journal of Parapsychology, 40*, 1–33.

Krippner, S., Saunders, D. T., Morgan, A., & Quan, A. (2019). Remote viewing of concealed target pictures under light and dark conditions. *Explore, 15*, 27–37. https://doi.org/10.1016/j.explore.2018.07.001

Kurdi, B., Lozano, S., & Banaji, M. R. (2017). Introducing the open affective standardized image set (OASIS). *Behavior Research Methods, 49*, 457–470. https://doi.org/10.3758/s13428-016-0715-3

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1995). *International Affective Picture System (IAPS): Technical manual and affective ratings*. Gainesville: University of Florida, Center for Research in Psychophysiology.

Lange, R. (2017). Rasch scaling and cumulative theory-building in consciousness research. *Psychology of Consciousness: Theory, Research and Practice, 4*, 135–160. https://doi.org/10.1037/cns0000118

Lange, R., & Houran, J. (2021). Replicable survey evidence for an "enchantment-psi" loop. *Journal of*

Transpersonal Psychology, 53, 140–156.

Lange, R. & Houran, J. (2024). Development of the Haunted People Syndrome Screener (HPSS). *Journal of Transpersonal Psychology*, *56*, 6–26.

Lange, R., Laythe, B., & Houran, J. (2023). Preregistered field test of an 'enchantment-psi' loop. *Journal of Parapsychology, 87*, 11–34. http://doi.org/10.30891/jopar.2023.01.03

Lange, R., Thalbourne, M.A., Houran, J., & Storm, L. (2000). The Revised Transliminality Scale: Reliability and validity data from a Rasch top-down purification procedure. *Consciousness and Cognition, 9*, 591–617. https://doi.org/10.1006/ccog.2000.0472

Lantz, N.D., Luke, W.L.W., & May, E.C. (1994). Target and sender dependencies in anomalous cognition experiments. *Journal of Parapsychology, 58*, 285–302.

Laythe, B., Houran, J., Lange, R., & Boussoffara, M.A. (2021). A 'multi-event sensor app' (MESA 3.0) for environmental studies of exceptional human experiences. *Australian Journal of Parapsychology, 21*, 128–162.

Laythe, B., Houran, J., & Ventola, A. (2018). A split-sample psychometric study of haunters. *Journal of the Society for Psychical Research, 82*, 193–218.

Laythe, B., & Roberts, N. (2022). A four element-themed self-selecting mobile application for ESP testing. *Journal of Scientific Exploration, 36*, 682–691. https://doi.org/10.31275/20222789

Li, L., Hew, K.F., & Du, J. (2024). Gamification enhances student intrinsic motivation, perceptions of autonomy and relatedness, but minimal impact on competency: A meta-analysis and systematic review. *Educational Technology Research and Development, 72*, 765–796. https://doi.org/10.1007/s11423-023-10337-7

Luhrmann, T.M. (1989). *Persuasions of the witch's craft: Ritual magic in contemporary England*. Harvard University Press.

MacDonald, D.A., Kuentzel, J.G., & Friedman, H.L. (1999a). A survey of measures of spiritual and transpersonal constructs: Part one – Research update. *Journal of Transpersonal Psychology, 31*, 137–154.

MacDonald, D.A., Kuentzel, J.G., & Friedman, H.L. (1999b). A survey of measures of spiritual and transpersonal constructs: Part two-additional instruments. *Journal of Transpersonal Psychology, 31*, 155–177.

MacDonald, D.A., LeClair, L., Holland, C.J., Alter, A., & Friedman, H.L. (1995). A survey of measures of transpersonal constructs. *Journal of Transpersonal Psychology, 27*, 171–235.

Marchewka, A., Żurawski, Ł., Jednoróg, K., & Grabowska, A. (2014). The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods, 46*, 596–610. https://doi.org/10.3758/s13428-013-0379-1

Mayer, G.A. (2024). A multi-method perspective on psi. *Journal of Scientific Exploration, 38*, 317–322. https://doi.org/10.31275/20243461

Mello, M.M., Goodman, S.N., & Finkelstein, J.A. (2020). Standardized Protocol Items Recommendations for Observational Studies (SPIROS): Protocol for a Delphi Study. *JMIR Research Protocols, 9*, Article e17864. https://doi.org/10.2196/17864

Milton, J., & Wiseman, R. (1999). Does psi exist? Lack of replication of an anomalous process of information transfer. *Psychological Bulletin, 125*, 387–391. https://doi.org/10.1037/0033-2909.125.4.387

Montgomery, R.A., & Packard, E. (1979–1998). *Choose your own adventure* [Book series]. Bantam Books.

Morris, R.L. (2001). Research methods in experimental parapsychology: Problems and prospects. *European Journal of Parapsychology, 16*, 8–18.

Morse, J. (2003). Principles of mixed methods and multimethod research design. In C. Teddlie & A. Tashakkori (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 189–208). SAGE Publications.

Mossbridge, J., Green, D., French, C.C., Pickering, A., & Abraham, D. (2025). Future dreams of electric sheep: Case study of a possibly precognitive lucid dreamer with AI scoring. *International Journal of Dream Research, 18*, 151–168. https://doi.org/10.11588/ijodr.2025.2.108750

O'Keeffe, C., Houran, J., Caputo, G.B., Katz, D.L., & Williamson, J. (2024). Magic flights or mind's eye? Competing content analyses of dimensional-slip narratives. *Journal of Transpersonal Psychology, 56*, 147–186.

OpenAI. (2025). ChatGPT-5 (Version 5) [Large language model]. https://chatgpt.com/.

Ortega, J.R., Gross, M., & Schooler, J. (2025). When life is but a dream: Does transliminality predict continuity of thought across the sleep-wake cycle? *Philosophy and the Mind Sciences, 6*, 1–21. https://doi.org/10.33735/phimisci.2025.10272

Parker, A., Grams, D., & Pettersson, C. (1998). Further variables relating to psi in the ganzfeld. *Journal of Parapsychology, 62*, 321–337.

Pooley, A.L. (2025). *Optimising methodological practices within psychology: A parapsychology case study* (Doctoral thesis, School of Philosophy, Psychology and Language Sciences, The University of Edinburgh). http://dx.doi.org/10.7488/era/6177

Rabeyron, T., & Watt, C. (2010). Paranormal experiences, mental health and mental boundaries, and psi. *Personality and Individual Differences, 48*, 487–492. https://doi.org/10.1016/j.paid.2009.11.029

Radin, D. (1997). *The conscious universe: The scientific truth of psychic phenomena*. HarperEdge.

Radin, D. (2006). *Entangled minds: Extrasensory experiences in a quantum reality*. Paraview/Harper.

Radin, D. (2025). Observer influence on quantum interference: Testing the von Neumann-Wigner consciousness-collapse theory. *Physics Essays, 38*, 64–80. https://doi.org/10.4006/0836-1398-38.1.64

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests* (Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B.D. Wright. University of Chicago Press.

Ritchie, S.J., Wiseman, R., & French, C.C. (2012). Failing the future: Three unsuccessful attempts to replicate Bem's 'retroactive facilitation of recall' effect. *PLoS One, 7*, Article e33423. https://doi.org/10.1371/journal.pone.0033423

Rock, A. J., Houran, J., Tressoldi, P. E., & Laythe, B. (2023). Is biological death final? Recomputing the *Drake-S* Equation for postmortem survival. *International Journal of Transpersonal Studies, 42*, 9–26. https://doi.org/10.24972/ijts.2023.42.1.9

Rosen, C., Park, S., Baxter, T., Tufano, M., & Giersch, A. (2023). Sensed presence, attenuated psychosis, and transliminality: At the threshold of consciousness. *Psychopathology, 56*, 359–370. https://doi.org/10.1159/000528572

Roxburgh, E. C., Vernon, D. & Schofield, M. B. (2024). Sensory processing sensitivity, transliminality, and boundary-thinness as predictors of anomalous experiences, beliefs, and abilities. *Current Psychology, 43*, 30098–30106. https://doi.org/10.1007/s12144-024-06619-9

Schooler, J. W., Baumgart, S., & Franklin, M. (2018). Entertaining without endorsing: The case for the scientific investigation of anomalous cognition. *Psychology of Consciousness: Theory, Research, and Practice, 5*, 63–77. https://doi.org/10.1037/cns0000151

Sheldrake, R. (1998). Experimenter effects in scientific research: How widely are they neglected? *Journal of Scientific Exploration, 12*, 73–78.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*, 1359–1366. https://doi.org/10.1177/0956797611417632

Smith, M. D. (2003). The role of the experimenter in parapsychological research. In J. E. Alcock, J. E. Burns, & A. Freeman (Eds.), *Psi wars: Getting to grips with the paranormal* (pp. 69–84). Imprint Academic.

Springer Nature Research Communities. (2021, October 15). No data! What to do when your article isn't founded on data (or reused others' data). *Springer Nature Communities*. https://go.nature.com/2XbwizE

Storm, L., & Tressoldi, P. E. (2017). Gathering in more sheep and goats: A meta-analysis of forced-choice sheep-goat ESP studies, 1994–2015. *Journal of the Society for Psychical Research, 81*, 79–107.

Storm, L., & Tressoldi, P. E. (2023). Assessing 36 years of the forced choice design in extra sensory perception research: A meta-analysis, 1987 to 2022. *Journal of Scientific Exploration, 37*, 517–535. https://doi.org/10.31275/20232967

Storm, L., Tressoldi, P. E., & Di Risio, L. (2012). Meta-analysis of ESP studies, 1987–2010: Assessing the success of the forced-choice design in parapsychology. *Journal of Parapsychology, 76*, 243–273.

Swami, V., Pietschnig, J., Stieger, S., Voracek, M., & Tran, U. S. (2024). Transliminality-Converging evidence of associations with and openness to experience and its facets. *Zeitschrift für Psychologie, 232*, 269–278. https://doi.org/10.1027/2151-2604/a000576

Targ, R. (2012). *The reality of ESP: A physicist's proof of psychic abilities*. Quest Books/Theosophical Publishing House.

Thalbourne, M. A. (2000). Transliminality and creativity. *Journal of Creative Behavior, 34*, 193–202. https://doi.org/10.1002/j.2162-6057.2000.tb01211.x

Thalbourne, M.A., & Houran, J. (2003). Transliminality as an index of the sheep-goat variable. *European Journal of Parapsychology, 18*, 3–14.

Thalbourne, M.A., & Maltby, J. (2008). Transliminality, thin boundaries, unusual experiences, and temporal lobe lability. *Personality and Individual Differences, 44*, 1617–1623. https://doi.org/10 .1016/j.paid.2008.01.022

Thalbourne, M.A., & Storm, L. (2012). Has the sheep-goat variable had its day? Testing transliminality as a psi predictor. *Australian Journal of Parapsychology, 12*, 69–80.

Thomas, D.M., Rycroft, E., Allan, K., & Sharif, S. (2026). New methodological directions for involving children in past life memories research. *Explore, 22*, Article 103283. https://doi.org/10.1016/j.explore.2025.103283

Tressoldi, P., & Katz, D.L. (2023). Remote viewing: A 1974–2022 systematic review and meta analysis. *Journal of Scientific Exploration, 37*, 467–489. https://doi.org/10.31275/20232931

Ventola, A., Houran, J., Laythe, B., Storm, L., Parra, A., Dixon, J., & Kruth, J.G. (2019). A transliminal 'disease' model of poltergeist 'agents.' *Journal of the Society for Psychical Research, 83*, 144–171.

Vernon, D. (2020). *Dark cognition: Evidence for psi and its implications for consciousness*. Routledge. https://doi.org/10.4324/9780429447488

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H.L.J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem, 2011. *Journal of Personality and Social Psychology, 100*, 426–432. https://doi.org/10.1037/a0022790

Wahbeh, H., Radin, D., Cannard, C., & Delorme, A. (2022). What if consciousness is not an emergent property of the brain? Observational and empirical challenges to materialistic models. *Frontiers in Psychology, 13*, Article 955594. https://doi.org/10.3389/fpsyg.2022.955594

Walleczek, J., von Stillfried, N., Schmidt, S., Wittmann, M., Kirmse, K.A., Moll, J., & Kekecs, Z. (2025). Metascientific replication project with the advanced meta-experimental protocol of the transparent psi project procedures for testing the precognitive effect claimed by Bem. *PLoS One, 20*, Article e0335330. https://doi.org/10.1371/journal.pone.0335330

Wang, C., Yu, L., Mo, Y., Wood, L.C., & Goon, C. (2022). Pareidolia in a built environment as a complex phenomenological ambiguous stimuli. I*nternational Journal of Environmental Research and Public Health, 19*, Article 5163. https://doi.org/10.3390/ijerph19095163

Watt, C.A. (1996). What makes a good psi target? Three studies of forced-choice ESP varying target emotionality and complexity. *Journal of Parapsychology, 60*, 25-41.

Watt, C., & Nagtegaal, M. (2004). Reporting of blind methods: An interdisciplinary survey. *Journal of the Society for Psychical Research, 68*, 105–114.

Wertz, F.J. (2025). Toward an inclusive methodology for psychology. *The Score, Division 5 Periodical, 47*, 2–6.

Westlund, E., & Stuart, E.A. (2017). The nonuse, misuse, and proper use of pilot studies in

experimental evaluation research. *American Journal of Evaluation, 38*, 246–261. https://doi.org/10.1177/1098214016651489

Williams, G. (2013). Psi and the problem of consciousness. *Journal of Mind and Behavior, 34*, 259–283.

Wiseman, R., & Schlitz, M. (1997). Experimenter effects and the remote detection of staring. *Journal of Parapsychology, 61*, 197–207.

Zemel, S. K., & Wahbeh, H. (2025). Code of ethics for mediums and trance channelers. *Explore, 21*, Article 103159. https://doi.org/10.1016/j.explore.2025.103159

Erweiterte deutsche Zusammenfassung

Entwicklung einer Dual-Mode-Anwendung für die Psi-Forschung

Wir stellen die Entwicklung und Pilotevaluierung des Immersive Psi Test (IPT) vor, einer Dual-Mode-Anwendung zur Erforschung anomaler Kognition. Die IPT integriert Forced-Choice-Messungen mit KI-gestützter narrativer Immersion, um das ökologische Engagement zu verbessern. Er wurde konzipiert, um eine seit langem bestehende methodologische Spannung in der Parapsychologie zwischen der Objektivität von Forced-Choice-Paradigmen und der Erfahrungsvielfalt von Free-Response-Protokollen wie Ganzfeld und Remote Viewing anzugehen. Durch die Einbettung strukturierter, objektiv bewerteter Antwortoptionen in eine spielerisch gestaltete narrative Umgebung in der zweiten Person, die den klassischen *Choose Your Own Adventure*-Büchern nachempfunden ist, versucht der IPT, die Stärken beider Ansätze zu vereinen. Die Anwendung wurde in Python unter Verwendung von Streamlit erstellt, auf GitHub gehostet und so konfiguriert, dass die Daten an einen sicheren Amazon S3-Bucket übermittelt werden. Das große Sprachmodell (LLM) ChatGPT-5 wurde verwendet, um 400 immersive narrative Passagen vorab zu generieren – acht für jeden der 50 fotografischen Reize in zwei experimentellen Bedingungen. Die Vorabgenerierung wurde gegenüber Live-Abfragen bevorzugt, um die Varianz aufgrund von LLM-Drift und Netzwerkinstabilität zu vermeiden.

Der IPT operationalisiert zwei unterschiedliche Indizes für mutmaßliches Psi. Explizites Psi wird anhand von fünf bildbasierten Forced-Choice-Durchgängen erfasst, in denen die Teilnehmenden angewiesen werden, mithilfe ihrer psychischen Intuition ein verdeckt vorausgewähltes Zielbild aus fünf Bildern zu identifizieren (Zufallswahrscheinlichkeit = 20 %). Implizites Psi wird anhand von zehn verdeckten narrativen Entscheidungspunkten erfasst, die zwischen den expliziten Durchgängen verstreut eingefügt sind. Dabei wählen die Teilnehmer eine Bewegungsrichtung – vorwärts, rückwärts, links oder rechts – als natürlichen Handlungsverlauf,

ohne zu wissen, dass diese Entscheidungen anhand eines versteckten, zufällig ausgewählten Zielbildes bewertet werden (Zufallswahrscheinlichkeit = 25 %). Diese Dual-Mode-Architektur ermöglicht die gleichzeitige Messung bewusst-intentionaler und unbewusst-spontaner psi-ähnlicher Effekte innerhalb einer einzigen, etwa 10-minütigen Sitzung.

Die Teilnehmenden wurden per Zufall entweder der „Enchanted"-Bedingung zugewiesen, die 25 Fotos mit hoher Valenz umfasste, die auf die Erzeugung psychologischer Absorption und Numinosität in sechs Gestaltdimensionen der Umgebung normiert waren, oder einer „Disenchanted"-Bedingung, die 25 Fotos mit niedriger Valenz von alltäglichen Umgebungen umfasste, zugeteilt. Die Stimuli stammten aus einer normierten Sammlung von Houran und Laythe (2026). Der Valenzunterschied zwischen den Bedingungen war groß und statistisch abgesichert (t[45,10] = 12,83, p < 0,001, d = 3,63). Randomisierungskontrollen mittels Chi-Quadrat-Anpassungsgütetests bestätigten, dass die Zielzuordnungen im Großen und Ganzen einer Gleichverteilung entsprachen, trotz geringfügiger Abweichungen in einem expliziten Versuch (Versuch 5, p = 0,023, V = 0,15) und bei zwei narrativen Variablen (Narrative 1b und 5b), die als technische Aspekte für zukünftige konfirmatorische Untersuchungen vermerkt sind.

In einer Pilotstudie mit einer Convenience-Stichprobe (N = 126), die über die CloudResearch-Plattform rekrutiert wurde, wiesen die Teilnehmenden ein breites Altersspektrum auf (M = 37,4 Jahre, SD = 12,6) mit ausgewogenem Geschlechterverhältnis und unterschiedlichem ethnischem Hintergrund. Die Teilnehmenden wurden nicht auf psychometrische Einschlusskriterien überprüft, jedoch wurden Daten zur Transliminalität (Revised Transliminality Scale; RTS) und zur Neigung zum Kontakt mit übersinnlichen Phänomenen (Haunted People Syndrome Screener; HPSS) erhoben, um die Funktionalität der Instrumente zu bewerten und vorläufige Korrelationen mit der Psi-Leistung zu untersuchen.

Die Ergebnisse zeigten einen deutlichen Unterschied zwischen den beiden Psi-Indizes. Die explizite Psi-Leistung übertraf weder in der Bedingung „Enchanted" (Trefferrate = 0,18, exakter Binomialtest p = 0,85, h = −0,06) noch in der Bedingung „Disenchanted" (Trefferrate = 0,19, exakter Binomialtest p = 0,64, h = −0,02) oder in der Gesamtstichprobe (Trefferrate = 0,19, exakter Binomialtest p = 0,83, h = −0,04) den Zufallserwartungswert. Im Gegensatz dazu erzielte die implizite Psi-Leistung in der „Enchanted"-Bedingung eine Trefferquote von 0,29 gegenüber einer Zufallserwartung von 0,25 und erreichte statistische Signifikanz bei einem einseitigen exakten Binomialtest (p = 0,016, Cohen's h = 0,09). Obwohl dieser Wert unter den üblichen Schwellenwerten für kleine Effekte (h = 0,20) liegt, ist er konsistent mit Effekten, die in Paradigmen impliziter Kognition und in früheren Forced-Choice-Psi-Metaanalysen beobachtet wurden. Weder die „Disenchanted"-Bedingung (Trefferquote = 0,24, exaktes p = 0,68) noch die kombinierte Stichprobe (Trefferquote = 0,27, exaktes p = 0,10) erreichten das Signifikanzniveau

für implizites Psi. Spearman-Korrelationen bestätigten den erwarteten moderaten Zusammenhang zwischen Transliminalität und Neigung zum Kontakt mit übersinnlichen Phänomenen ($\rho = 0{,}52$, $p < 0{,}001$) und replizierten damit frühere Befunde. Allerdings sagte keine der beiden Messgrößen in dieser unselektierten Stichprobe explizites Psi, implizites Psi oder die Reaktionslatenz voraus, und die beiden Psi-Indizes korrelierten untereinander nicht ($\rho = -0{,}01$), was darauf hindeutet, dass sie funktional unabhängige Prozesse widerspiegeln.

Die Distanzierung zwischen nullwertigem explizitem Psi und über dem Zufallsniveau liegendem implizitem Psi in der „Enchanted"-Bedingung steht theoretisch im Einklang mit Ganzfeld-Befunden, die darauf hindeuten, dass entspannte, passive Aufmerksamkeitszustände – und nicht etwa bewusstes, angestrengtes Streben – mit einer höheren Psi-Leistung assoziiert sind. Sie deckt sich auch mit weiterführenden theoretischen Diskussionen über ein Anstrengungsparadoxon in der anomalen Kognition. Die verdeckte Gestaltung narrativer Entscheidungspunkte könnte intuitive Reaktionen ohne Leistungsdruck ermöglichen ohne die Leistungbeeinträchtigung, die den Erfolg bewusst gesteuerter Versuche niederhält. Diese Interpretation ist jedoch vorläufig: Überzufällige Leistungen bei verdeckten Entscheidungen könnten auch unbewusste Antwortverzerrungen, Erwartungseffekte oder unterschiedliche Aufmerksamkeitsbelastungen widerspiegeln, und das vorliegende Versuchsdesign kann nicht zwischen diesen Alternativen unterscheiden. Der Begriff „implizites Psi" wird daher deskriptiv und nicht als mechanistische Behauptung verwendet.

Diese Ergebnisse sollten als Proof-of-Concept-Ergebnisse einer unselektierten Stichprobe und nicht als konfirmatorischer Psi-Nachweis betrachtet werden. Die Studie wurde nicht vorab registriert und zielte nicht an das Profil hoher Transliminalität und hoher Neigung zum Kontakt mit übersinnlichen Phänomenen ab, von dem theoretisch angenommen wird, dass es die Leistung optimiert. Zukünftige Forschungen sollten vorab registrierte Designs mit geschichteter Stichprobenziehung von begabten Probanden, verbesserten Block-Randomisierungsdiagnostiken, intrasubjektiven Crossover-Vergleichen expliziter und impliziter Formate, psychophysiologischen Messungen zur Untersuchung zugrunde liegender Mechanismen sowie Sensitivitätsanalysen zur Quantifizierung des Einflusses geringfügiger Randomisierungsunregelmäßigkeiten verwenden. Das IPT-Framework steht qualifizierten Forschern auf schriftliche Anfrage zur Verfügung und ist darauf ausgelegt, transparente, reproduzierbare und kumulative Wissenschaft zu anomaler Kognition innerhalb einer skalierbaren, Open-Science-orientierten Architektur zu unterstützen.

Schlüsselbegriffe: immersives Testen, Instrumentalisierung, Liminalität, narrative Einbindung, Psi

Journal of Anomalistics
Zeitschrift für Anomalistik

www.anomalistik.de www.igpp.de

Ghosthunting-Gruppen und Ghosthunter in Deutschland

Ergebnisse zweier Online-Umfragen[1]

Gerhard Mayer,[a*] Sonja Nowara[b]

Zusammenfassung – Wir haben zwei Online-Umfragen mit deutschen Ghosthunting-Gruppen (GHGs) und einzelnen Ghosthuntern durchgeführt. Der erste Fragebogen richtete sich an Teams, der zweite an einzelne Mitglieder. Mit dieser Studie wollten wir (1) ein besseres Verständnis der aktuellen GHG-Szene in Deutschland gewinnen und Informationen zur Zusammensetzung der Gruppen, den von ihnen vertretenen Untersuchungsansätzen und den außergewöhnlichen Erfahrungen (AgEs), die sie im Kontext des Ghosthuntings gemacht hatten, erhalten; und (2) wollten wir etwas über die Menschen erfahren, die sich diesem zeit- und ressourcenintensiven Hobby widmen. Wir entwickelten dazu zwei Fragebögen. In den zweiten Fragebogen nahmen wir zwei etablierte Skalen auf, um außergewöhnliche Erfahrungen (Fach, PAGE II) und paranormale Überzeugungen (Schofield et al., BitSS) zu erheben. Wir erhielten 21 ausgefüllte Fragebögen von GHGs und 39 von Einzelpersonen.

Ergebnisse Studie 1: Wir fanden große Ähnlichkeiten zwischen den GHGs hinsichtlich der Wahl der Untersuchungsorte (Locations), der Untersuchungsmethoden, des Einsatzes technischer Ausrüstung und der aufgezeichneten Anomalien. Soziale Medien ersetzen mittlerweile weitgehend traditionelle Websites und lokale Fernsehberichte. Die Teams unterscheiden sich jedoch in ihren Dokumentationspraktiken

(a) Institut für Grenzgebiete der Psychologie und Psychohygiene (IGPP), Freiburg i.Br.

(b) Unabhängige Forscherin Bonndorf

*Korrespondierender Autor: Gerhard Mayer mayer@igpp.de

1 Dies ist eine deutsche Übersetzung des Originalartikels „Ghost Hunting Groups and Ghost Hunters in Germany: Results of Two Online Surveys", der im *Journal of the Society for Psychical Research, 90*(2), 65–100, erschienen ist. https://doi.org/10.63499/kta4es60

und im Umgang mit den gesammelten Daten. Gruppengrößen, Gründungsdaten und die Anzahl der Untersuchungen pro Jahr variieren stark.

Ergebnisse Studie 2: Wie erwartet, berichten Ghosthunter von mehr AgEs als eine studentische Referenzstichprobe, und sie haben stärkere paranormale Überzeugungen (BitSS). 85 % nehmen einen Zusammenhang zwischen paranormalen Phänomenen und Verstorbenen an, obwohl sich 30 % als Atheisten oder Agnostiker bezeichnen – was darauf hindeutet, dass der Glaube an die Geister Verstorbener nicht unbedingt mit dem Glauben an einen Gott verbunden ist. Für Ghosthunter sind AgEs in der Regel nicht beunruhigend, sondern wecken Neugier und den Wunsch nach weiteren Erfahrungen dieser Art. Fast zwei Drittel gaben an, dass frühere AgEs ausschlaggebend für den Einstieg in das Ghosthunting waren.

Neben dem Unterhaltungsaspekt spielt die Möglichkeit eines „echten" Kontakts mit dem Übernatürlichen für viele Ghosthunter eine wichtige Rolle – etwas, das in den Medienangeboten zu diesem Thema nicht zu finden ist. Die Motivation der Ghosthunter reicht von wissenschaftlichem Interesse bis hin zum Wunsch, die eigenen Überzeugungen bestätigt oder in Frage gestellt zu bekommen.

Schlüsselbegriffe: Ghosthunting-Gruppen, außergewöhnliche Erfahrungen, paranormale Überzeugungen, technische Ausrüstung, paranormale Untersuchungen

Einleitung

Ghosthunting (GH) ist eine Freizeitaktivität, die seit der Jahrtausendwende in Deutschland zunehmend an Beliebtheit gewonnen hat (z. B. Mayer, 2013). Ghosthunter sind Menschen, die sich in kleinen Gruppen zusammenschließen und in ihrer Freizeit Orte aufsuchen, an denen es angeblich spukt. Manche Ghosthunting-Gruppen (GHGs) führen auf Anfrage von betroffenen Personen auch „paranormale Untersuchungen" (PUs) von Spukfällen in Privathaushalten durch. Das Hauptziel des GH besteht darin, paranormale Phänomene an den betreffenden Locations mithilfe verschiedener technischer Geräte wie Audio- und Videorekordern, Vollspektrum- und Infrarotkameras, Messgeräten für elektromagnetische Felder (EMF-Messgeräte) und Temperatur sowie Radiofrequenz-Bewegungsmeldern (REM-Pods) zu erfassen und aufzuzeichnen (z. B. Auerbach, 2004; Wilson, 2005).[2] Der Einsatz technischer Geräte gilt als Merkmal einer wissenschaftlichen Haltung. Er reflektiert die Vorstellung von vielen Ghosthuntern, dass „Geister" physikalische Spuren erzeugen bzw. hinterlassen, die gemessen werden können (Mayer, 2010; Potts, 2004). Die Messdaten und Aufzeichnungen werden nach den Feldunter-

2 Mittlerweile gibt es eine Vielzahl von Ratgebern zu den technischen Aspekten der Geisterjagd, unter anderem auch das Buch *Ghost Hunting for Dummies* (Bagans, 2020). Die jüngeren Veröffentlichungen liefern allerdings kaum neue Informationen.

suchungen zuhause analysiert, und die Ergebnisse werden der Öffentlichkeit häufig in Form von Forschungsberichten auf den eigenen Websites, auf YouTube und auf Facebook zugänglich gemacht.

Die meisten der bislang veröffentlichten Studien zum Phänomen des GH stützen sich auf die Analyse der Selbstdarstellungen von GHGs im Internet (z. B. Hill, 2017; Mayer, 2013; Potts, 2004) und werden teilweise durch Interviews ergänzt (Bartoschek & Waschkau, 2013; Mayer, 2010). Der amerikanische Literaturwissenschaftler Alan Brown führte strukturierte Telefoninterviews mit 43 Gründern oder Leitern von GHGs durch und fragte nach Motiven, Methoden, nach der Gruppenzusammensetzung, dem Hintergrund der Mitglieder und vielem mehr (Brown, 2006). Er stellte bei ihnen eine Reihe von Gemeinsamkeiten fest, beispielsweise hinsichtlich ihrer Mission, der verwendeten technischen Ausrüstung, ihrer Erfahrungen mit den sozialen Reaktionen ihres Umfelds und – überraschenderweise – ihrer Ablehnung der Fernsehsendung *Ghost Hunters*, die für die Entwicklung der GH-Szene von großer Bedeutung war (Mayer, 2013). Den wichtigsten Unterschied zwischen den einzelnen Gruppen sah Brown darin, ob sie regelmäßig mediale Personen oder Wünschelruten als „metaphysische Methoden" einsetzten, um zusätzliche Informationen zu erhalten, oder ob sie sich auf rein technische Messgeräte beschränkten, die „harte" Daten lieferten. Mehr als die Hälfte der Gruppen nutzte sowohl „metaphysische" als auch „wissenschaftliche" Methoden (Brown, 2006, S. 345–348).

Für die Geologin Sharon Hill stand die Frage nach dem wissenschaftlichen Charakter der Aktivitäten von paranormalen Laienforschungsgruppen im Mittelpunkt ihrer Studie (Hill, 2017), während sich der Anthropologe Ehler Voss auf die quasi-religiösen Aspekte bei GHGs in seiner Feldforschung in Kalifornien konzentrierte (Voss, 2021). In einer sehr aufschlussreichen Studie über GHGs führte die Kulturanthropologin Michele Hanks Interviews mit Ghosthuntern und Anbietern von Ghosthunting-Tourismus-Angeboten (z. B. „Geisterwanderungen" in Städten) durch und lernte durch partizipative Beobachtung die Arbeit von vier GHGs mit einer Mitgliederzahl zwischen zwei und 25 kennen (Hanks, 2015, 2016b, 2016a). Sie erörterte die wichtige Rolle des Zweifels für Ghosthunter, der deren Verbindung zur Wissenschaft gewährleisten soll: Wer an vermeintlich paranormalen Phänomenen zweifelt und ihnen kritisch gegenübersteht, handelt rational. Das Zweifeln zeigt an, dass man nicht leichtgläubig und leicht beeinflussbar ist.

Sonja Nowara stellte ihre Erkenntnisse als langjährige Insiderin der GH-Szene in einem Artikel vor. Neben einem kurzen historischen Überblick und einer Darstellung der von GHGs verwendeten gängigen Methodik beschrieb sie die unterschiedlichen Selbstdarstellungen der Gruppen sowie neuere Entwicklungen und Veränderungen in der deutschen GH-Szene – hauptsächlich aus einer emischen Perspektive (Nowara & Mayer, 2024).

Unseres Wissens wurden bisher noch keine systematischen Erhebungen zu persönlichen Motivationen, Überzeugungen und anderen persönlichen Aspekten von Ghosthuntern durchgeführt. Das Ziel unseres Forschungsprojekts war es, einen Überblick über die aktuelle Szene in Deutschland, über die Zusammensetzung der GHGs, den spezifischen Ansatz bei PUs und die außergewöhnlichen Erfahrungen (AgEs) zu gewinnen, die während der PUs erlebt werden; außerdem wollten wir Informationen über die Menschen gewinnen, die dieser zeit- und ressourcenintensiven Freizeitbeschäftigung nachgehen. Wir wollten wissen, welche „Art von Menschen" sich davon angezogen fühlt und welche individuellen Motive die einzelnen Ghosthunter haben. Außerdem wollten wir Informationen darüber gewinnen, welche Art von AgEs individuell erlebt worden sind und wie ausgeprägt der Glaube an paranormale und übernatürliche Phänomene ist. Zu diesem Zweck erstellten wir zwei Fragebögen, von denen sich der erste an ganze GH-Teams und der zweite an Einzelpersonen richtete. Aufgrund des Mangels an vergleichbaren Studien war diese zweiteilige Untersuchung rein explorativ. Sie wurde von der lokalen Ethikkommission des Instituts für Grenzgebiete der Psychologie und Psychohygiene (IGPP-2023-02) genehmigt.

Umfrage 1: Ghosthunting-Gruppen in Deutschland

Methode

Teilehmende

Der erste Fragebogen richtete sich an ganze Teams. Wir erhielten insgesamt 22 Datensätze. Zwei Mitglieder einer GHG füllten den Fragebogen getrennt voneinander für dieses Team aus, sodass sich die insgesamt gewonnenen Informationen auf 21 GHGs beziehen. Für die statistische Gruppenauswertung verwendeten wir die Daten der Teamleiterin dieser GHG, da sie auch deren Gründerin ist.

Materialien

Der für die GHGs entwickelte Fragebogen enthielt Fragen zum geografischen Standort des Teams, zu den Untersuchungsmethoden bei den PUs, zum Umgang mit Medien (Fernsehen, Printmedien, soziale Medien) usw. Er wurde nicht anonymisiert, da keine personenbezogenen Daten abgefragt wurden, sondern lediglich Daten, die prinzipiell öffentlich auf den Websites der jeweiligen Gruppen zugänglich sind. Der Fragebogen umfasste 37 Items, von denen einige

Items Freitextfelder für detailliertere Beschreibungen enthielten. Der Fragebogen war in fünf Blöcke gegliedert: (1) allgemeine Angaben oder Informationen wie die Anzahl der Mitglieder, die Anzahl der durchgeführten PUs, die zur Veröffentlichung der Fälle genutzten Internetplattformen usw.; (2) Fragen zur Durchführung von PUs (Locations, technische Ausrüstung, Zusammenarbeit mit öffentlichen Medien); (3) Fragen zur Dokumentation und Auswertung von PUs; (4) Fragen zu paranormalen Phänomenen und AgEs während PUs; und (5) Bezug zur akademischen Wissenschaft. Der Fragebogen enthielt zudem einen einleitenden Abschnitt, in dem wir uns und unsere Forschungsfrage vorstellten. Wir nutzten das Online-Fragebogen-Tool LimeSurvey.

Vorgehensweise

Der Fragebogen wurde an aktive GHGs in Deutschland verschickt, deren Kontaktdaten verfügbar waren. Dies war bei 35 der damals 58 uns namentlich bekannten aktiven Gruppen der Fall. Der Aufruf wurde zudem in zwei speziellen GH-WhatsApp-Gruppen („Ghosthunter-Treffen" und „Blabla") sowie auf Facebook veröffentlicht. Die Datenerhebung erfolgte zwischen Oktober und Dezember 2023.

Datenanalyse

Für statistische Berechnungen verwendeten wir SPSS (Version 28). Die meisten Statistiken sind rein deskriptiver Natur. Für Gruppenvergleiche verwendeten wir den Kruskal-Wallis-Test für ordinal skalierte Messwerte. Aufgrund des rein explorativen Charakters der Korrelationsstatistiken nahmen wir keine Korrekturen für multiples Testen vor. Wir führten zweiseitige Signifikanztests durch und wählten ein Signifikanzniveau von 0,05 (Spearman's Rho).

Ergebnisse

Allgemeine Fragen oder Aussagen

Alter und Größe der Gruppen: Das Durchschnittsalter der GHGs betrug 7,7 Jahre (*SD* = 5,95). Die Spanne reichte von 0 Jahren, d. h. diese Gruppen wurden im Jahr der Erhebung gegründet (3 Gruppen = 14,3 %), bis zu 19 Jahren (1 Gruppe = 4,8 %). Die Hälfte der Teams war nicht älter als sieben Jahre, und ein Viertel war nicht älter als 2,5 Jahre. Dies ist ein Indikator dafür, dass die Gruppen tendenziell nicht besonders lange bestehen. Allerdings existierte etwa ein Viertel der Teams bereits seit mehr als 12 Jahren. Die Gruppen hatten durchschnittlich drei Mitglieder (*SD* = 1,8), mit mindestens einer Person (Einzelforscher) und höchstens acht Personen. Hunde, mit denen regelmäßig gearbeitet wurde, wurden nicht als Mitglieder gezählt.

Anzahl der paranormalen Untersuchungen

Zur Beantwortung der Frage nach der Anzahl der durchgeführten PUs wurden Freitextfelder verwendet, in die ein entsprechender Eintrag vorgenommen werden sollte. Dies ermöglichte den Befragten bei Bedarf eine Differenzierung, führte jedoch auch zu weniger eindeutigen Antworten. Bei ungefähren Angaben wurden die niedrigeren Werte herangezogen; bei Angabe einer Spanne wurde der Mittelwert (abgerundet) verwendet.

Im Durchschnitt führten die Gruppen sieben bis acht PUs pro Jahr durch (*Median [Mdn]* = 7; *Interquartilsabstand [IQR]* = 5–11), wobei die Spanne zwischen zwei und 24 lag. Der *Mdn* der Gesamtzahl der seit Gründung der Gruppen durchgeführten PUs beträgt 40 (*IQR* = 20–95) bei einer Spanne von sieben bis 120 PUs. Diese Zahlen deuten auf eine hohe Varianz sowohl hinsichtlich des Engagements (Anzahl der PUs pro Jahr) als auch hinsichtlich der Beständigkeit oder Kontinuität der Gruppen hin.

Mehr als die Hälfte der Gruppen hatte maximal vier PUs in Privathaushalten durchgeführt (*Mdn* = 4; *IQR* = 1–15), was als eine besondere Art von PU angesehen werden kann (siehe unten). In einer Gruppe mit einer sehr hohen Anzahl an PUs betrafen jedoch etwa die Hälfte der Fälle Untersuchungen in Privathaushalten.

GHGs und die Öffentlichkeit

In den Anfangsjahren des modernen GH war die Berichterstattung über die PU auf den eigenen Websites der wichtigste Weg, um die Aufmerksamkeit der Öffentlichkeit oder anderer Gruppen zu gewinnen. Von den an der Umfrage teilnehmenden Teams verfügen 38 % über eine eigene Website, und 29 % nutzten diese, um ihre PU-Berichte zu veröffentlichen. Die Bedeutung einer eigenen Website für diesen Zweck hat inzwischen jedoch abgenommen. Zum Zeitpunkt der Umfrage war die am häufigsten genutzte Plattform zur Veröffentlichung von Untersuchungsberichten Facebook (86 %), gefolgt von YouTube (67 %) und Instagram (57 %). Andere Plattformen werden von 29 % genutzt.

Wir haben auf einer Fünf-Punkte-Skala (von „völlig unwichtig“ [=1] bis „sehr wichtig“ [=5]) gefragt, wie wichtig Zuschauerzahlen, Seitenaufrufe und der allgemeine Bekanntheitsgrad für das Team sind. Für die meisten Gruppen (81 %) sind diese Aspekte „völlig unwichtig“ oder „nicht sehr wichtig“. Nur ein Team hält diesen Aspekt für „wichtig“. Keines der Teams stufte ihn als „sehr wichtig“ ein (siehe Tabelle 1).

Tabelle 1

Bedeutung der öffentlichen Wahrnehmung: „Wie wichtig sind Zuschauerzahlen, Seitenaufrufe und der allgemeine Bekanntheitsgrad für das Team?“

	Häufigkeit	Prozent	Kumulierte Prozentsätze
Völlig unwichtig (=1)	8	38,1	38,1
Nicht sehr wichtig (=2)	9	42,9	81,0
Etwas wichtig (=3)	3	14,3	95,2
Wichtig (=4)	1	4,8	100
Sehr wichtig (=5)	0	0	
Gesamt	21	100	

Die Beziehungen der Teams zu anderen GHGs

Wir baten die Teams, ihre Beziehung zu anderen GHGs auf einer Fünf-Punkte-Skala von „überwiegend gut“ bis „überwiegend problematisch“ einzuschätzen. Zwei Drittel der GHGs (14 Gruppen) beschrieben die Beziehung zwischen ihrem Team und anderen Teams als überwiegend gut, weitere 14,3 % (3 Gruppen) als allgemein gut, weitere drei Teams als „teilweise gut/teilweise problematisch“, und nur ein Team beschrieb die Beziehung zu anderen GHGs als „allgemein problematisch“. Die Kategorie „überwiegend problematisch“ war nicht belegt.

Vorbilder

Wir fragten die Teams nach Vorbildern als Orientierung für die PUs. Interessanterweise scheinen Fernsehserien wie die einflussreiche amerikanische Sendung Ghost Hunters keine Rolle mehr zu spielen. Offenbar hat sich ihr Einfluss auf Ausrüstung und Technologie so sehr etabliert, dass ein Rückgriff auf die ursprünglichen Inspirationsquellen nicht mehr nötig ist. Die Websites anderer GHGs sind für fast 43 % ein wichtiger Einflussfaktor. Es scheint jedoch noch andere Vorbilder zu geben, da rund 76 % bei dieser Frage „Sonstiges“ ankreuzten (siehe Tabelle 2).

Tabelle 2

Vorbilder für GHGs

	Häufigkeit	Prozent
Amerikanische Fernsehserien (z. B. *Ghost Hunters*)	0	0
Andere Fernsehserien (z. B. *Geisterakten*)	2	9,5
Fernsehdokumentationen	3	14,3
Websites / Präsentationen anderer Teams	9	42,9
Sonstiges	16	76,2

Anmerkung: Mehrfachantworten waren möglich

Durchführung von PUs

Location der PUs

Der wichtigste Faktor bei der Wahl der Location für eine PU besteht darin, dass sie mit einem bestimmten historischen Ereignis verbunden ist. Fünfundneunzig Prozent der Gruppen führten PUs an solchen, für die Öffentlichkeit zugänglichen Orten durch[3] und 76 % auch an Orten, die nicht für die Öffentlichkeit zugänglich sind.[4] Locations ohne konkreten historischen Bezug, wie beispielsweise Friedhöfe, sind für GHGs ebenfalls von Interesse: 67 % führten PUs an öffentlich zugänglichen Orten dieser Art durch und 48 % auch an Orten, die nicht öffentlich zugänglich sind.

Die Untersuchung von Spukphänomenen in Privathaushalten kann als Sonderfall von PUs an nicht öffentlich zugänglichen Locations betrachtet werden. Es handelt sich dabei jedoch um eine besondere Art von PU, die von einigen Gruppen kritisch gesehen wird. An einer öffentlich zugänglichen Location oder einem Lost Place, an dem ortsgebundene Spukphänomene auftreten, kann niemand während einer PU psychischen Schaden erleiden (mit Ausnahme der Ghosthunter selbst). Dies ist bei Spuk-Fällen in Privathaushalten nicht der Fall. Hier haben wir es in der Regel mit personenbezogenem Spuk zu tun, also mit Phänomenen, die meist mit komplexen psychodynamischen Faktoren zusammenhän-

3 Typische Beispiele sind Burgen oder andere historisch bedeutsame Gebäude, die als Touristenattraktionen besichtigt werden können.

4 Typische Beispiele sind verlassene Kliniken oder andere „Lost Places" auf Privatgrundstücken, zu denen normalerweise kein Zutritt gewährt wird.

gen (Fach, 2022, 2024). Nur 38 % der Gruppen untersuchten Spuk- oder Poltergeist-Fälle in Privathaushalten. Dies ist immer noch eine relativ hohe Zahl, wenn man bedenkt, dass solche Untersuchungen in psychodynamische Prozesse eingreifen, die oft professioneller psychotherapeutischer Begleitung bedürfen.

62 % der Teams gaben an, bei der Auswahl einer Location keine weiteren spezifischen Kriterien zu beachten, während 38 % dies tun. Pragmatische Aspekte wie z. B. Einsturzgefahr der Gebäude, bekannte Störquellen (z. B. stark befahrene Straßen) sowie vorliegende (Dreh-) Genehmigungen der Eigentümer wurden (in Kommentarfeldern) als relevante Faktoren genannt.

Vorbereitung der PUs

Alle Teams recherchieren vor einer PU zur Geschichte der Location. Es gibt unterschiedliche Meinungen darüber, wie gründlich man im Vorfeld historische Recherchen durchführen sollte, da sie durch erzeugte Erwartungshaltungen die Wahrnehmung beeinflussen können. Dies gilt auch für das Rezipieren von Untersuchungsberichten anderer GHGs, die die Location bereits zuvor untersucht hatten. Einige Teams sind jedoch der Ansicht, dass die Kenntnis sogenannter Hotspots bei der Auswahl und Installation der Ausrüstung hilfreich und zeitsparend ist. Einige Gruppen legen besonderen Wert auf gezielte Untersuchungen zu möglichen Störquellen (z. B. Funkmasten). In manchen Fällen verfügt nur der Teamleiter über Vorabinformationen zur Location, um die Unvoreingenommenheit der Teammitglieder so weit wie möglich zu gewährleisten.

Technische Ausrüstung und sonstige Hilfsmittel

Audio- und Videoaufzeichnungsgeräte sowie EMF-Messgeräte gehören zur Grundausstattung der Teams. Alle Teams arbeiten mit solchen Geräten, wobei in einem Fall, in dem kein Audioaufnahmegerät verwendet wird, die Audiospur der Videokamera genutzt wird. Aufgrund der spärlichen Lichtverhältnisse an der Location (alle Lichtquellen werden während der Untersuchung ausgeschaltet) werden Videoaufnahmen im Infrarot- oder Ultraviolettbereich (Vollspektrum) gemacht. In den meisten Fällen werden die Videokameras und Audiogeräte auf der Grundlage vorheriger Recherchen an (vermuteten) „Hotspots“ installiert. Zusätzlich zur in allen Teams verfügbaren Standardausrüstung kommen weitere Geräte zum Einsatz, z. B. REM-Pods (= Radiofrequenz-Bewegungsmelder, die einen Ton auslösen, wenn sich ein Objekt durch das von den Geräten erzeugte elektromagnetische Feld bewegt), Spirit Boxes und Portals (modifizierte Radioempfänger zur Aufnahme elektrischer Stimmphänomene [EVPs]), Kinects (bewegungsempfindliche Eingabegeräte, die ursprünglich von Microsoft für die Xbox

360 entwickelt wurden), Parascopes (zur Visualisierung von Veränderungen im umgebenden elektromagnetischen Feld) und andere Geräte (siehe Tabelle 3).[5]

Tabelle 3

Von GHGs verwendete technische Ausrüstung

	Prozent
EMF-Meter (inkl. K2, Mel Meter)	100
Videoaufzeichnungsgeräte/Kameras (inkl. Infrarot-, Vollspektrumkameras)	100
Tonaufzeichnungsgeräte	95
Rem-Pod	76
Spirit Box	57
Kinect	43
Handy-Apps	33
Portal	29
Parascope	24
Sonstiges	76

Anmerkung: Mehrfachantworten waren möglich

Teilweise sind GHGs auch der Ansicht, dass es „lebende Detektoren" gibt, nämlich medial begabte Personen oder auch Tiere, denen eine besondere Sensibilität für die Anwesenheit unsichtbarer Wesen zugeschrieben wird und die neben technischen Messgeräten zur Erkennung von Anomalien eingesetzt werden (Mayer & Schetsche, 2019, S. 96–97).

Etwas mehr als die Hälfte der Teams (11 von 21) arbeitet gelegentlich mit einem Medium zusammen, und in fast der Hälfte der Teams (10 von 21) gab es Personen, die glauben, selbst über mediale Fähigkeiten zu verfügen.

Zusammenarbeit mit den öffentlichen Medien

Die Frage der Zusammenarbeit mit dem Fernsehen oder der Presse während der PUs beschäftigt die GHGs von Beginn an. Die Anwesenheit von Fernsehteams lenkt die öffentliche Aufmerksamkeit auf die Gruppen, was von einigen aus verschiedenen Gründen geschätzt wird

5 Eine kurze technische Erklärung solcher Geräte findet sich unter https://www.darkwhimsicalart.com/blogs/news/ghost-hunting-equipment-gadgets-that-work-or-just-expensive-toys?srsltid=AfmBOorQBaUcOvz5eXI_HCs7SRzOd8NIgUUbWlmT3nMtWDVG6gCa60fb

– schließlich sind die Mitglieder der amerikanischen GHG „The Atlantic Paranormal Society" (TAPS) dank der Serie *Ghost Hunters* zu Medienstars geworden (Mayer, 2013). Neben solchen persönlichen Aspekten kann die Medienpräsenz auch dazu führen, dass die Teams Informationen über mögliche Locations vermittelt bekommen („Nur so können wir den Menschen zeigen, dass es uns gibt und dass sie sich an uns wenden können") oder leichter die Erlaubnis für eine PU erhalten. Doch ist es ein zweischneidiges Schwert. „Presse und TV (haben) ganz bestimmte Vorstellungen und Wünsche", was dem Forschungsteam Freiheit und Spontaneität nimmt. Gewünscht werden sensationelle Bilder oder Filmmaterial. Zudem besteht immer die Gefahr, dass man „durch derartige Medien (…) verspottet und ins Lächerliche gezogen" wird (alle Zitate stammen aus Kommentaren der Teams zur Frage nach der Zusammenarbeit mit den Medien).

57 % der GHGs lehnten solche Kooperationen während der PUs ab. Von den 43 %, die angaben, dies zuzulassen, liegen uns keine Informationen darüber vor, wie oft dies geschieht. Ein Kommentar zu dieser Frage lautete: „Prinzipiell nein. Ab und an gibt es Ausnahmen (…) Deswegen ‚ja', aber es werden 90 % der Anfragen abgelehnt." Manche machen die Zusammenarbeit auch von der Art der Anfrage abhängig, beispielsweise davon, ob sie von regionalen Medien oder vom öffentlich-rechtlichen Fernsehen stammt. Insgesamt besteht die Tendenz, sensationslüsterne Darstellungen abzulehnen. Dies wird durch den folgenden Kommentar veranschaulicht:

> Aus langjähriger Erfahrung und zahlreichen TV-Zusammenarbeiten kann ich inzwischen sagen, dass eine solche Zusammenarbeit unbefriedigend ist und dem Thema nicht gerecht wird. Gesucht wird Grusel & Action, es versteht sich von selbst, dass das so nicht geliefert werden kann und ein falscher Eindruck in der Öffentlichkeit vorprogrammiert ist.

Der Ablauf

Der Ablauf einer PU ist bei allen GHGs sehr ähnlich. Nach einer Begehung der Location (um ein Gefühl für den Ort zu bekommen, ihn auf mögliche Störquellen zu prüfen, den Standort zu begehen, um vorhandene Gefahren zu identifizieren und zu markieren sowie Hotspots zu ermitteln) findet eine Beratung statt, gefolgt vom Aufbau der Ausrüstung, in einigen Fällen kombiniert mit einer „Basismessung" z. B. von Temperatur, EMF-Strahlung und, in Einzelfällen, der Luftqualität. Einige der Teams beginnen mit einer Meditation, oft gefolgt von einer „passiven Sitzung" oder einem „passiven Sit-in", d. h., Umgebungsgeräusche werden bewusst wahrgenommen und identifiziert (sehen/fühlen/riechen/hören). Darauf folgt in der Regel eine „aktive Sitzung", d. h., es werden Fragen gestellt, um „Kontakt" herzustellen. Dieser Ablauf wird

an verschiedenen Orten an der Location wiederholt. In den meisten Fällen wird dies durch Kameras überwacht. Die Befragungen und Messungen dauern oft bis in die frühen Morgenstunden und erstrecken sich über drei bis sechs Stunden, wobei es Abweichungen in beide Richtungen gibt. Manchmal teilt sich das Team auf, und bestimmte Bereiche der Location werden in einer bestimmten Gruppenzusammensetzung untersucht (z. B. nur männliche oder nur weibliche Teammitglieder). Am Ende wird die verwendete Ausrüstung abgebaut.

Dokumentation und Datenauswertung

Bei der Dokumentation von PUs sind zwei Aspekte zu berücksichtigen: (1) die Erfassung potenzieller Anomalien wie Klopfgeräusche, Schatten oder die Messausschläge bzw. Reaktionen der in Tabelle 3 aufgeführten Geräte, und (2) die Dokumentation der subjektiven Wahrnehmungen der Ghosthunter, des Ablaufs selbst sowie der allgemeinen Bedingungen während des PU. Die meisten Teams nutzen Video- und Audioaufnahmen sowie Fotos zu Dokumentationszwecken. Etwas mehr als die Hälfte der GHGs (57 %) fertigt zusätzlich zur Datengewinnung mittels technischer Aufnahmegeräte schriftliche Aufzeichnungen an. Zwei GHGs zeichnen die gesamten PUs mit technischen Aufnahmegeräten auf, während die meisten nur einzelne Sitzungen (z. B. während einer Phase des Fragens oder während einer Phase der Verwendung bestimmter GH-Geräte) als Teile der gesamten PU aufzeichnen. Unmittelbar nach den Sitzungen werden die Teilnehmenden zu ihren (subjektiven) Wahrnehmungen befragt. In vielen Teams sprechen die Teilnehmenden auch direkt während der einzelnen Sitzungen über ihre Wahrnehmungen, sodass diese Äußerungen auf den Video- und Audioaufnahmen festgehalten werden. Grundlegende Informationen wie Ort, Datum, Namen der Anwesenden und Wetterdaten werden ebenfalls technisch und/oder schriftlich festgehalten. Besondere Vorkommnisse oder Umstände am Ort werden dokumentiert, etwa auffällige Hintergrundgeräusche (Autos, Flugzeuge), Störfaktoren (andere Personen, Tiere) oder selbst verursachte Geräusche (z. B. Gähnen, Magenknurren), die bei der Analyse zu Fehlinterpretationen führen könnten. Alle vor Ort angefertigten Aufzeichnungen und sonstigen Dokumentationen fließen in die anschließende Analyse ein. Zwei Teams verwenden einen selbst erstellten Fragebogen zur Dokumentation, der neben den grundlegenden Informationen (Ortsname, Datum, Uhrzeit) die einzelnen Geräte und den Namen der Person, die das jeweilige Gerät auswertet, sowie die Art der möglichen Anomalie erfasst.

Die Daten werden auf PCs übertragen und analysiert. Jede Aufzeichnung wird durchgesehen bzw. durchgehört, etwaige Anomalien werden mit einem Zeitcode vermerkt und dann im Kontext mit anderen Aufzeichnungen vom selben Ort verglichen (manchmal befinden sich

mehrere Kameras und Audiogeräte in einem einzigen Raum oder begrenzten Bereich). Dies dient dazu, festzustellen, ob eine bestimmte Anomalie auf mehreren Geräten zu sehen oder zu hören ist. Ein Beispiel hierfür wäre, wenn auf einer Aufzeichnung eine visuelle Anomalie zu sehen ist und zur gleichen Zeit ein oder mehrere Teammitglieder ungewöhnliche (körperliche) Wahrnehmungen (z. B. auf der Audioaufzeichnung) berichtet haben. Dies erleichtert die Einschätzung, ob es sich tatsächlich um Anomalien oder um konventionell erklärbare Phänomene handelt.

Die Ergebnisse der Auswertung werden schriftlich und/oder als Videodokumentation festgehalten und in einigen Fällen in sozialen Medien veröffentlicht.

Hinsichtlich der Art der ausgewerteten Daten (Video- und Audioaufnahmen, Fotos) gibt es keine Unterschiede zwischen den Teams. Dies gilt auch für die eher subjektiven Daten. Trotz der großen Affinität der Ghosthunter zur Technik und dem damit verbundenen Wunsch nach physikalischer Messbarkeit der Phänomene werden subjektive Wahrnehmungen, Empfindungen und Erlebnisse (z. B. körperliche Empfindungen, das Sehen von Schatten, Berührungsempfindungen) von allen Teams mit Ausnahme eines dokumentiert und ausgewertet.

Die Antworten auf die Frage, welche Teammitglieder die Daten auswerten, ergeben kein einheitliches Bild. Alle Varianten sind gleichermaßen verbreitet und hängen nicht von der Größe des Teams ab. In 33 % der Fälle findet zunächst eine Auswertung durch Spezialisten im Team statt, gefolgt von einer gemeinsamen Auswertung. 29 % der Teams gaben an, dass die Gruppenmitglieder die Auswertung gemeinsam durchführen. In 24 % der Teams erfolgt dies ausschließlich durch Spezialisten, und in 14 % der Teams werten die Mitglieder die Daten teilweise gemeinsam aus. Persönliche Vorlieben und Interessen einzelner Teammitglieder spielen hier eine entscheidende Rolle. Für manche mag die Durchführung und das Erleben einer PU vor Ort am wichtigsten sein, während der Auswertungsteil als eher langweilig empfunden und gerne anderen Teammitgliedern überlassen wird. Zwei Drittel der Teams ziehen gelegentlich externe Experten zur Auswertung hinzu oder holen Gutachten ein.

71,5 % der Teams speichern und bewahren alle Rohdaten auf einem Speichermedium oder in der Cloud auf. Knapp 10 % speichern auf solche Weise nur die Rohdaten interessanter Teile oder Abschnitte der PUs. In den verbleibenden 19 % der Fälle werden die Rohdaten nach der Analyse oder Verarbeitung größtenteils gelöscht. 76 % der GHGs geben an, dass sie die bearbeiteten Daten vollständig sichern und speichern, während 19 % nur die interessantesten Teile speichern.

76 % der GHGs stellen ihre bearbeiteten Daten öffentlich zur Verfügung (soziale Netzwerke, Websites usw.), 24 % auch die Rohdaten. 64 % veröffentlichen zudem ihre historischen

Recherchen zu den Locations, und 57 % liefern Kontextinformationen (z. B. Weg- und Umgebungsbeschreibungen).

Potenziell paranormale Phänomene und AgEs während PUs

Da das zentrale Ziel beim GH darin besteht, paranormale Phänomene zu erleben und aufzuzeichnen, sind entsprechende Berichte und Aussagen zu erwarten. Wir waren daran interessiert, welche solcher Phänomene erlebt wurden und ob sie dokumentiert werden konnten. Die folgende Tabelle 4 zeigt eine Liste der berichteten Phänomene und den Anteil der GHGs, die solche Phänomen jeweils erlebt haben.

Tabelle 4

Wahrgenommene paranormale Phänomene, die während PUs erlebt wurden

	Prozent
Akustische Phänomene (Mimikry-Geräusche, z. B. Klopfen, Stimmen)	100
Visuelle Phänomene (z. B. Schatten)	95
Haptische Wahrnehmungen, Berührungen	90
Das Gefühl, dass jemand oder etwas anwesend ist / ein Geist / eine Energie	81
Plötzliche Temperaturänderungen	71
Vorübergehende gesundheitliche Beeinträchtigungen (z. B. Schmerzempfindungen)	67
Besondere Geruchswahrnehmungen	67
Gefühl der Bedrohung	52
Gefühl, von einer fremden Wesenheit beeinflusst zu werden	29
Sonstiges	24

Anmerkung: Mehrfachantworten waren möglich

Einige der genannten Phänomene betreffen die Ebene der Emotionen und der inneren Wahrnehmung und sind nicht objektiv messbar. Daher erfolgt die Dokumentation in diesen Fällen in Form eines schriftlichen Protokolls oder durch das Ausfüllen eines zu diesem Zweck erstellten Fragebogens. Andere Phänomene werden nicht während der PU selbst wahrgenommen, sondern erst später bei der Auswertung der aufgezeichneten Daten entdeckt. Tabelle 5 enthält eine Liste solcher dokumentierten Phänomene mit den entsprechenden Prozentsätzen.

Tabelle 5

Aufgezeichnete oder dokumentierte paranormale Phänomene während der PUs

	Prozent
Elektrische Stimmphänomene (EVPs)	100
Akustische Phänomene (z. B. Mimikry-Geräusche, Klopfen)	95
Reaktionen und Messaufzeichnungen von Geräten, die bei der Geisterjagd eingesetzt werden	90
Anomalien auf Fotos	81
Ungewöhnliche Reaktionen einzelner Teammitglieder	38
Deutliche Temperaturschwankungen	14
Sonstiges	24

Anmerkung: Mehrfachantworten waren möglich

Alle aufgeführten Kategorien wahrgenommener Phänomene, mit Ausnahme des Gefühls, von einer fremden Wesenheit beeinflusst zu werden, werden von mindestens der Hälfte der Gruppen berichtet (Tabelle 4). Das Setting des GH scheint ungewöhnliche akustische, visuelle oder haptische Wahrnehmungen zu befördern. Dies gilt auch für das Gefühl der Anwesenheit von jemandem oder etwas (einem Geist / einer Energie). Vier Fünftel der Gruppen berichteten über Letzteres.

Auch die dokumentierten, als paranormal eingestuften Phänomene bieten kaum Überraschungen (Tabelle 5). Alle Gruppen fanden auf ihren Aufnahmen elektrische Stimmphänomene (EVPs) sowie andere akustische Phänomene. Reaktionen von „Geisterdetektoren" und anderen spezifischen Geräten für GH konnten auf Video aufgezeichnet werden. 81 % der Gruppen berichteten von Anomalien auf Fotografien. Bemerkenswert ist höchstens, dass Temperaturschwankunegn, die für klassische Erzählungen von Geistergeschichten eher typisch sind, nur selten objektiv gemessen werden konnten.

Wir befragten die Teams zu ihren eindrucksvollsten paranormalen Erlebnissen im Zusammenhang mit PUs. Aufgrund der Vielfalt der berichteten Phänomene, die erst in Kombination oft ein starkes Gefühl subjektiver Evidenz auslösten, ist der Versuch, sie auseinanderzudividieren und trennscharfen Kategorien zuzuordnen, wenig sinnvoll, wie aus den folgenden Beispielen ersichtlich ist. Ein Team berichtete:

> RSPK-Phänomen in Schottland. Zwei von fünf Personen (ich selbst war eine davon) hörten und spürten Hammerschläge, als kommen sie aus dem Nebenraum. Ein im Fenster platziertes Gerät wurde, nach Aussage des zufällig beiwohnenden und nicht zum Unter-

> suchungsteam gehörenden Hausbesitzers, angehoben und nach vorne fallen gelassen. Erst beim Aufprall des Gerätes auf den Boden endeten die Hammerschläge.

Ein anderer Bericht lautet:

> Ich wurde im Garten von einem Lehmklumpen beworfen, zur gleichen Zeit nahmen wir eine EVP auf, in der eine Frau sagte: wirf ihn weg. Ich sah mal eine Planchette aus einer Metallschüssel ca. 3 Meter im hohen Bogen fliegen. Da diese Schüssel ca. 1,5 m neben mir stand und ich alle sich im Raum befundenen Menschen sehen konnte, kann ich mir das bis heute nicht erklären.

Während die beiden oben genannten Beispiele akustische Phänomene und sich auf unerklärliche Weise bewegende Objekte kombinieren, betrifft das nächste Beispiel eine bemerkenswerte fotografische Anomalie.

> Beim Besuch der Burg Hohenlimburg im Rahmen der Dreharbeiten für ein Format des WDR[6] konnten wir zwei Fotos machen welche mich bis heute beeindrucken. Die Fotos einer weißen Frau welche auch namentlich bekannt war/ist. Die Klarheit der Fotos und die Tatsache das die Fotos in einem fast leeren Raum entstanden sind haben mich bis heute nicht losgelassen. Noch während der Untersuchung haben wir versucht diese Fotos zu reproduzieren um auszuschließen das es nicht doch eine rationale Erklärung dafür gibt, aber ohne Erfolg. Für mich als objektiven Skeptiker sind diese Fotos ein Beweis dafür das es eben doch Geister gibt!

Das letzte hier vorgestellte Beispiel handelt von körperlichen Empfindungen, die während einer paranormalen Untersuchung auftraten.

> Stich (mit einem Finger?) in den Rücken und nach meiner Aufforderung, es sein zu lassen, ein sanftes Streicheln über meinen Arm und das Ergreifen meiner Hand. Es fühlte sich an wie die Hand eines Kindes – wobei weder hinter mir, noch neben mir genügend Platz für ein Kind gewesen wäre.

Trotz der Vielfalt der beschriebenen Phänomene haben fast alle gemeinsam, dass sie die physikalische Ebene betreffen. Dazu gehören beispielsweise unerklärliche Bewegungen von Gegenständen oder das Auftreten von Geräuschen, für die herkömmliche Ursachen ausgeschlossen wurden.

Der Ausschluss konventioneller Erklärungen ist in der GH von zentraler Bedeutung. Die Teams finden für etwa drei Viertel der Phänomene konventionelle Erklärungen (M = 68,5

6 WDR = Westdeutscher Rundfunk – Anm. d. Red.

%, *SD* = 27 %, *Mdn* = 75 %). Wir haben um kurze Beschreibungen des Ausschlussverfahrens bezüglich konventioneller Erklärungen für vermeintlich paranormale Phänomene gebeten. Fast alle Teams gaben an, dass sie zunächst Ausschlussmethoden und Messungen anwenden, um nach natürlichen Ursachen zu suchen, und dabei auch Experten hinzuziehen. So werden beispielsweise EMF-Messungen durchgeführt, um zu überprüfen, ob starke elektromagnetische Felder vorhanden sind, ob sich in der Nähe Strom- oder Mobilfunkmasten befinden, ob es natürliche Störquellen gibt, die sich in den aufgezeichneten Daten widerspiegeln könnten, usw. Werden beispielsweise Anomalien auf Fotos festgestellt, so werden oft Tests vor Ort zur Reproduzierbarkeit der Phänomene durchgeführt, indem man sie nachzustellen und zu rekonstruieren versucht. Bei der Analyse von Daten, die mit verschiedenen Geräten aufgezeichnet wurden, spielt deren zeitliche Synchronisation eine wichtige Rolle. Damit soll festgestellt werden, ob eine potenzielle Anomalie nur auf einem Gerät oder in irgendeiner Form auf mehreren Geräten nachzuweisen ist.

Zwar gibt es innerhalb der GHGs nicht immer eine einheitliche Bewertung der Phänomene, doch insgesamt ist die Übereinstimmung in den meisten Fällen recht groß. Über 80 % geben an, dass es am Ende der Analyse nur „selten" oder „gelegentlich" zu widersprüchlichen Meinungen kommt. Nur zwei Teams (10 %) geben an, dass dies häufig der Fall ist.

GH und akademische Wissenschaft

Drei abschließende Fragen im Gruppenfragebogen betrafen wissenschaftliche Theorien und die Bereitschaft zur Zusammenarbeit mit akademischen Forschern. Auf die Frage, ob sich das Team auch mit wissenschaftlichen Erklärungsmodellen oder Theorien zu diesem Thema befasse, gaben 62 % an, dies gelegentlich zu tun. Vier GHGs (19 %) tun dies intensiv, und weitere vier GHGs beschäftigen sich nicht damit. Namen, die in den begleitenden Kommentaren genannt werden, gehören zum Bereich der Transkommunikation (z. B. Ernst Senkowski) und der Parapsychologie (Hans Bender, Walter v. Lucadou, Alan Gould, Steve Parsons und andere), wobei oft nicht zwischen populärwissenschaftlicher und akademischer Literatur unterschieden wird. Die Mehrheit der Teams (71 %) wäre sehr an einem engeren wissenschaftlichen Austausch mit akademischen Forschern interessiert und auch bereit, bestimmte Anforderungen für eine Zusammenarbeit mit der Wissenschaft zu erfüllen (z. B. das Ausfüllen eines Fragebogens vor und nach der PU). Vierundzwanzig Prozent (= fünf Teams) wären „eher" interessiert; ein Team gab an, „nicht sehr" interessiert zu sein. Vier Teams gaben an, dass sie bestimmte Bedingungen für eine Zusammenarbeit „wahrscheinlich" erfüllen würden, während die übrigen zwei mit „eher nicht" antworteten.

Diskussion 1

Es lassen sich in vielerlei Hinsicht Ähnlichkeiten zwischen den GHGs feststellen. Dies betrifft die Auswahl der Locations, die Art des Vorgehens, den Einsatz technischer Ausrüstung, die erlebten und dokumentierten Anomalien, aber auch die Tatsache, dass Fernsehserien wie *Ghost Hunters* oder *Most Haunted* nicht als Vorbilder angesehen werden. GHGs suchen Anleitung auf den Websites anderer Teams oder in anderen Quellen, was auf einen Wandel in der Szene hindeutet. Die Präsenz in sozialen Medien hat GHG-Websites und Auftritte bei lokalen Fernsehsendern weitgehend ersetzt. Angesichts der zunehmenden Kommerzialisierung von Teilen der GHG-Szene (Nowara & Mayer, 2024) sind die aus unserem Fragebogen gewonnenen Daten hinsichtlich der geringen Bedeutung von Zuschauerzahlen und öffentlicher Aufmerksamkeit überraschend. Ein weiteres gemeinsames Merkmal ist die Zurückhaltung der meisten GHGs, Untersuchungen in Privathäusern durchzuführen. Dies ist aus der Perspektive professioneller Berater für Menschen, die AgEs erlebt haben, positiv zu bewerten.

Hinsichtlich der Gruppengröße, des Alters der Gruppe und der Anzahl der pro Jahr durchgeführten PUs ist eine große Varianz zu beobachten. Dies ist angesichts des erheblichen Zeit- und Finanzaufwands, den das GH erfordert, verständlich.

Die Analyse und Interpretation von (Mess-)Daten sowie die Überprüfung verschiedener Erklärungshypothesen mittels Ausschlussverfahren sind gängige Ansätze des wissenschaftlichen Erkenntnisgewinns. Deshalb ist es verständlich, dass viele GHGs ihren Ansatz als wissenschaftlich betrachten und dass ihre Mitglieder sich selbst als skeptisch bezeichnen. GH ist jedoch eine Freizeitbeschäftigung, und nur sehr wenige Ghosthunter dürften eine Ausbildung in wissenschaftlicher Methodik erhalten haben, obwohl der Anteil derjenigen mit einem Fachhochschul- oder Universitätsabschluss unter den Teilnehmenden der Einzelinterview-Studie mit 38,5 % vergleichsweise hoch ist (siehe Ergebnisse von Umfrage 2). Grundsätzlich besteht eine hohe Bereitschaft zur Zusammenarbeit mit der akademischen Wissenschaft. Die Teams haben jedoch unterschiedliche Vorstellungen davon, wie die PUs dokumentiert werden und was genau mit den gesammelten Daten zu geschehen hat.

Einige Einschränkungen der Untersuchung sollten erwähnt werden. Da der Fragebogen nicht anonym war, könnten die Antworten einen erheblichen Bias im Sinne einer positiven Selbstdarstellung enthalten (z. B. die Beziehung zu anderen Gruppen, die überwiegend als gut beschrieben wird; oder dass Medienberichterstattung und Zuschauerzahlen für das Team nicht besonders wichtig seien). Es ist auch nicht bekannt, inwieweit die Daten jeweils eine gemeinsam erörterte Meinung des Teams widerspiegeln oder ob es sich um die individuelle Einschätzung des Teamleiters handelt. Zumindest hinsichtlich des letzten Punktes erhielten wir einen Hin-

weis auf die Unterschiede innerhalb einer GHG, da wir zwei Fragebögen erhielten, von denen einer von deren Teamleiter und der andere von einem weiteren Teammitglied (nur teilweise ausgefüllt) ausgefüllt wurde. Die meisten Unterschiede zwischen diesen beiden Datensätzen, etwa diejenigen bezüglich der Speicherung oder Veröffentlichung von Rohdaten, lassen sich auf unzureichende Kommunikation und unterschiedliche Informationsniveaus innerhalb der Gruppe zurückführen. Divergierende Antworten auf die Frage, was die beeindruckendste paranormale Erfahrung im Zusammenhang mit PUs war, zeigen deutlich, wie unterschiedlich solche Erfahrungen auf individueller Ebene wirken. Ein bemerkenswerter Unterschied zwischen den Aussagen des Teamleiters und des anderen Teammitglieds zeigte sich bei der Schätzung des Anteils der aufgezeichneten vermeintlich paranormalen Phänomene, für die das Team schließlich eine konventionelle Erklärung gefunden hat. Die Bewertung erfolgte mithilfe eines Schiebereglers im Bereich von 0 % bis 100 %. Der Teamleiter gab an, dass 25 % dieser Phänomene letztendlich konventionell erklärt werden konnten, während das andere Teammitglied den Anteil auf 75 % schätzte. Darüber hinaus nannten die beiden leicht unterschiedliche Häufigkeiten dafür, wie oft Teammitglieder am Ende der Analyse widersprüchliche Meinungen über die Natur der Phänomene hatten. Während der Teamleiter angab, dass dies „selten" der Fall sei, gab das andere Teammitglied an, dass dies „gelegentlich" vorkomme. Diese Unterschiede vermitteln den Eindruck, dass der Teamleiter weniger skeptisch als das andere Teammitglied ist. Ob dies allerdings eher auf die gewünschte Außendarstellung durch den Teamleiter zurückzuführen ist oder auf individuelle Unterschiede in der subjektiven Bewertung der Evidenz, lässt sich aus den Daten nicht ableiten.

Umfrage 2: Deutsche Ghosthunter

Methode

Teilnehmende

Der zweite Fragebogen richtete sich an einzelne Mitglieder der GHGs. Wir erhielten 39 Datensätze. Davon waren 38 vollständig, während bei einem die Werte auf der *Belief in the Supernatural Scale* (BitSS) fehlten. Die Rücklaufquote lässt sich nur grob schätzen, da wir nicht wissen, wie viele Mitglieder den Link zum Fragebogen erhalten hatten. Bei einer durchschnittlichen Gruppengröße von drei Personen (siehe Untersuchung 1) könnte sie bei etwa einem Drittel liegen. Das Durchschnittsalter der Teilnehmenden betrug 45 Jahre (SD = 9,3; Mdn = 46). 56 % waren weiblich, 44 % männlich, 0 % divers.

Die Mehrheit (36 %) fühlte sich dem Christentum zugehörig. Allerdings sind in der Stichprobe auch 21 % Atheisten und 9 % Agnostiker vertreten (siehe Tabelle 6).

Tabelle 6

Religion / Spiritualität

	Prozent
Christentum (katholisch / protestantisch)	36
Christentum (andere Formen)	0
Buddhismus	3
Islam	0
Sonstige	21
Unabhängige Spiritualität/Religiosität	23
Atheist („Ich glaube an nichts Göttliches")	21
Agnostiker („Man kann nicht entscheiden, ob Gott existiert oder nicht")	8

Anmerkung: Mehrfachantworten waren möglich

Der Familienstand in unserer Stichprobe entsprach in etwa dem Durchschnitt der deutschen Bevölkerung.[7] Etwa ein Viertel der Teilnehmenden (25,6 %) war ledig, etwas mehr als die Hälfte (51,3 %) war verheiratet, 20,5 % waren geschieden, und eine Person (2,6 %) war verwitwet. 72 % der Teilnehmenden hatten einen festen Lebenspartner.

Das Bild hinsichtlich der Bildungsabschlüsse ähnelt dem der Gesamtbevölkerung. 38,5 % hatten einen Fach-/Hochschulabschluss, 33,3 % einen Realschulabschluss und 28,2 % einen Hauptschulabschluss; niemand war ohne Schulabschluss.[8]

89,7 % der Teilnehmenden waren zum Zeitpunkt der Umfrage erwerbstätig. Dies ist ein höherer Wert als in der Gesamtbevölkerung, deren Erwerbsquote im Mai 2023 (Personen im Alter zwischen 25 und 74 Jahren) bei 68,1 % lag.[9] Drei Personen (7,7 %) waren Hausfrauen oder

7 www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsstand/Tabellen/familienstand-jahre-5.html

Die dort für 2023 angegebenen Zahlen mussten korrigiert werden, da sie sich auch auf Personen beziehen, die noch nicht volljährig sind.

8 www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Bildungsstand/Tabellen/bildungsabschluss.html

9 www.destatis.de/DE/Themen/Arbeit/Arbeitsmarkt/Erwerbslosigkeit/Tabellen/erwerbslose.html#129240

Hausmänner, und eine Person wählte „Sonstiges“. Die hohe Erwerbsquote ist verständlich, da GH ein kostspieliges Hobby ist.

Materialien

Zusätzlich zu den von uns selbst erstellten Fragebogenteilen verwendeten wir in diesem Fragebogen zwei bereits vorhandene Messinstrumente: den Fragebogen zur Phänomenologie außergewöhnlicher Erfahrungen (PAGE II) (Fach, 2018) und die Belief in the Supernatural Scale (BitSS; Schofield et al., 2018) in einer deutschen Übersetzung.[10] Der komplette Fragebogen bestand aus vier Abschnitten: (1) soziodemografische Daten (7 Fragen); (2) die Motivation und Einstellung gegenüber dem GH, einschließlich der Rolle früherer AgEs (5 Fragen); (3) Fragen zu AgEs im Allgemeinen (PAGE II; 20 Fragen); (4) Fragen zu übernatürlichen oder paranormalen Überzeugungen (BitSS; 44 Fragen). Er enthielt zudem einen einleitenden Abschnitt, in dem wir uns und unsere Forschungsfrage vorstellten.

Der PAGE II ist ein Messinstrument, das nach außergewöhnlichen Erfahrungen beispielsweise mit paranormalen Phänomenen wie Erscheinungen, Telepathie, Hellsehen, Vorahnungen und präkognitiven Träumen, aber auch nach Phänomenen wie seltsamen Wahrnehmungen (z. B. das Hören innerer Stimmen), Kognitionen (z. B. Gedankenintrusionen) und Automatismen (z. B. spontane und unkontrollierte Körperbewegungen) fragt. Die Autoren erstellten ein Klassifizierungsschema für AgEs, das diese in vier Grundkategorien auf zwei Dimensionen einteilt: (1) externale Phänomene – internale Phänomene und (2) Koinzidenzphänomene – Dissoziationsphänomene sowie in sechs typische AgE-Formenkreise (Spuk / Erscheinungen, internale Präsenz / Beeinflussung, außersinnliche Wahrnehmung, sinnvolle Fügung, Automatismen / Mediumismus, externale Präsenz / Nachtmahr) (Belz & Fach, 2015; Fach, 2022, 2024). Die Items sind den Grundkategorien zugeordnet und messen die Häufigkeit dieser Erfahrungen auf einer fünfstufigen Skala von „nie“ (= 0) bis „oft“ (= 4). Die Aktivität des GH selbst ließ zwar eine erhöhte Häufigkeit bestimmter AgEs vermuten, z. B. externale Präsenzen und Erscheinungen, aber wir wollten generell etwas über die Häufigkeit von AgEs auch außerhalb des GH erfahren.

Schofield et al. (2018) entwickelten die BitSS, um einige Probleme mit der weit verbreiteten Revised Paranormal Belief Scale (rPBS; Tobacyk, 2004) zu vermeiden. Unserer Meinung nach unterscheidet sie besser zwischen religiösen, übernatürlichen und paranormalen Überzeugungen. Die Skala weist eine Fünf-Faktoren-Struktur auf, bestehend aus „mentalen und psychologischen Phänomenen“, „religiösen Überzeugungen“, „Psychokinese“ (psychische Beeinflussung von Materie), „übernatürlichen Wesen“ und „allgemeinen paranormalen Wahrnehmungen“.

10 Die deutsche Übersetzung der BitSS wurde durch Mayer und Fuhrmann (2022) erstellt.

Die 44 Items messen den Grad der Zustimmung zu den Aussagen, die von „stimme überhaupt nicht zu" (= 1) bis „stimme voll und ganz zu" (= 7) reichen. Wir verwendeten diese Skala zur Untersuchung des Zusammenhangs der GH-Aktivität mit bestimmten persönlichen Glaubensüberzeugungen. Obwohl die Fragebogendaten keine Rückschlüsse darauf zulassen, ob AgEs paranormale Überzeugungen fördern oder ob umgekehrt der Glaube an Geister die Wahrscheinlichkeit erhöht, solche Phänomene wahrzunehmen (experiential vs. cultural source hypothesis; vgl. Hufford, 1982), versprachen die Daten interessante Erkenntnisse über den Zusammenhang zwischen den Werten in den Subskalen. GH an sich deutet auf eine spiritistische Weltanschauung hin, doch könnte man je nach dem Grad der Säkularisierung einer Gesellschaft kulturelle Unterschiede vermuten (Mayer, 2010, 2013). Durch die Anwendung des BitSS wollten wir zuverlässige Daten zu diesem Aspekt gewinnen.

Vorgehensweise

Der Fragebogen wurde gemeinsam mit dem Gruppenfragebogen an aktive GHGs in Deutschland verschickt. Die Teamleiter sollten den zweiten, individuellen Fragebogen an die Teammitglieder weiterleiten. Der Aufruf wurde zudem in zwei Ghosthunter-WhatsApp-Gruppen („Ghosthunter-Treffen" und „Blabla") sowie auf Facebook veröffentlicht. Die Datenerhebung erfolgte zwischen Oktober und Dezember 2023. Wir nutzten das Online-Fragebogen-Tool LimeSurvey. Der Fragebogen wurde gemäß den in Deutschland geltenden Datenschutzrichtlinien anonymisiert. Die Teilnehmenden gaben ihre Einwilligung nach Aufklärung und hatten die Möglichkeit, Kommentare abzugeben.

Datenanalyse

Für die statistischen Berechnungen verwendeten wir SPSS (Version 28). Für Gruppenvergleiche nutzten wir den Kruskal-Wallis-Test, da unsere Messinstrumente ordinal skaliert sind. Aufgrund des rein explorativen Charakters der Korrelationsstatistiken nahmen wir keine Korrekturen für multiple Tests vor. Wir führten zweiseitige Signifikanztests durch und einigten uns auf ein Signifikanzniveau von 0,05 (Spearman's Rho).

Ergebnisse

Motivation und Einstellung zum GH

Die Faszination des GH: Ein für fast alle Teilnehmenden faszinierender Aspekt des GH ist der Kontakt zum Paranormalen oder Übernatürlichen (95 %), gefolgt von der spezifischen

Geschichte des Ortes, an dem die PU durchgeführt wird (74 %). Ein weiterer wichtiger Faktor ist die Analyse der Daten. Auch diese übt auf drei Viertel der Teilnehmenden (74 %) eine besondere Faszination aus. Obwohl die Datenanalyse mühsam ist – beispielsweise das stundenlange Anhören oder Anschauen von aufgezeichnetem Audio- und Videomaterial –, scheint dieser Prozess dennoch eine große Anziehungskraft zu besitzen, vergleichbar vielleicht mit der Faszination des Schatzsuchers, der in der durchsiebten Erde etwas Wertvolles zu entdecken hofft. Die Faszination am Umgang mit technischen Geräten wurde von der Hälfte der Teilnehmenden genannt, das Gruppenerlebnis von nur 38 %, und die Aspekte der Öffentlichkeitsarbeit und -wirkung (Online-Präsentation und Popularität) spielten eine untergeordnete Rolle (siehe Tabelle 7).

Tabelle 7

Faszination am Ghosthunting

	Prozent
Der mögliche Kontakt zum Paranormalen / Übernatürlichen	95
Die Geschichte der Location, an denen PUs durchgeführt werden	74
Die Auswertung (Dokumentation, Analyse, Diskussion)	74
Die Locations, an denen PUs durchgeführt werdern	64
Der Umgang mit den in den PUs verwendeten technischen Geräten	51
Gruppenerlebnis	38
Die Präsentation der Ergebnisse auf der Website oder in sozialen Netzwerken	15
Die Medienwirksamkeit bzw. der Bekanntheitsgrad, der mit dem Veröffentlichen von Dokumentationen / filmischen Präsentationen erreicht werden kann	5

Anmerkung: Mehrfachantworten waren möglich

Erklärungen für paranormale Phänomene während der PUs: Wir boten drei Aussagen an, bei denen die Teilnehmenden auf einer 5-Punkte-Skala von „Ich stimme überhaupt nicht zu“ (= 1) bis „Ich stimme voll und ganz zu“ (= 5) ihren Grad der Zustimmung angeben mussten (siehe Tabelle 8): (1) Ich sehe sie als Zeichen von Verstorbenen, Geistern oder anderen Wesenheiten („spiritistisch“); (2) sie hängen mit den Mitgliedern des Teams zusammen („Team“); (3) sie hängen mit der Geschichte der Location zusammen („Location“).

Fast 85 % der Teilnehmenden halten die Erklärung 1 („spiritistisch“) für teilweise oder vollständig richtig. Drei Personen (7,7 %) können oder wollen sich nicht entscheiden, und weitere

drei lehnen diese Aussage teilweise oder vollständig ab. Fast 90 % sind davon überzeugt, dass die Phänomene mit der Geschichte der Location zusammenhängen („Location"). Drei Personen „wissen es nicht", und eine lehnt diese Erklärung vollständig ab. Mehr als die Hälfte der Teilnehmenden (53,9 %) stimmt der Erklärung „Team" teilweise oder vollständig zu. Ein Drittel weiß es nicht, und knapp 13 % lehnen diese Erklärung teilweise oder vollständig ab.

Tabelle 8

Grundlegende Erklärung oder Theorie zu paranormalen Phänomenen

	Ich stimme überhaupt nicht zu	Ich stimme teilweise nicht zu	Ich weiß es nicht	Ich stimme teilweise zu	Ich stimme voll und ganz zu
			Prozent		
Ich sehe sie als Zeichen von Verstorbenen, Geistern oder anderen Wesenheiten („spiritistisch")	2,6	5,1	7,7	43,6	41,0
Sie hängen mit den Mitgliedern des Teams zusammen („Team")	7,7	5,1	33,3	46,2	7,7
Sie hängen mit der Geschichte der Location zusammen („Location")	2,6	0	7,7%	79,5	10,3

Anmerkung: Mehrfachantworten waren möglich

Skeptische Haltung: Eine Frage betraf die Selbsteinschätzung der eigenen Skepsis gegenüber solchen außergewöhnlichen Phänomenen und Wahrnehmungen. Die Aufforderung lautete: *Bitte geben Sie mit dem Schieberegler auf einer Skala von 0 („überhaupt nicht skeptisch") bis 100 („extrem skeptisch") an, wie skeptisch gegenüber Wahrnehmungen bzw. Behauptungen von paranormalen oder übernatürlichen Phänomene Sie sich empfinden.* Der Mittelwert liegt bei 67,5 auf einer Skala von 0–100 (*SD* = 20,8; *Mdn* = 70) auf der Seite des skeptischen Pols.

Ghosthunter und außergewöhnliche Erfahrungen (AgEs)

Häufig diskutiert wird die Frage nach dem Zusammenhang zwischen kulturell bedingten Überzeugungen, individuellen Überzeugungen und AgEs. Nach der *Cultural Source Hypothesis* sind es kulturell überlieferte Vorstellungen und Erzählungen, die den Glauben an das Para-

normale oder Übernatürliche fördern und dann zu entsprechenden Erfahrungen führen. Die *Experiential Source Hypothesis* geht hingegen davon aus, dass gemachte Erfahrungen die Basis für entsprechende Glaubensvorstellungen bilden (Hufford, 1982). Diese beiden Hypothesen schließen sich nicht wechselseitig aus. Sie erklären unterschiedliche Aspekte der Beziehung zwischen direkten Erfahrungen, kommunizierten Erfahrungen, Überzeugungen und Erzählungen.

Wir befragten die Teilnehmenden zu ihren AgEs, bevor sie das GH als Hobby wählten, und zur Rolle, die diese AgE für die Zuwendung zum GH gespielt hatten. Etwa 64 % der Teilnehmenden (n = 25) hatten bereits vor dem Beginn des GH paranormale oder „übernatürliche" Erfahrungen gemacht. Ein weiteres Viertel (25,6 %; n = 10) gab an, dass sie glaubten, solche Erfahrungen gemacht zu haben, und nur vier Personen (10,3 %) hatten eindeutig keine vorherigen AgEs gehabt. 61,5% Prozent der Teilnehmenden gaben an, dass diese früheren AgEs eine entscheidende Rolle bei ihrer Entscheidung gespielt hätten, mit GH zu beginnen; für 28,2 % war dies nicht der Fall (siehe Tabelle 9).

Tabelle 9

AgEs vor der Aufnahme von GH als Hobby

	Ja, ich bin mir sicher	**Wahrscheinlich ja**	**Nein**
		Prozent	
Haben Sie paranormale oder übernatürliche Erfahrungen gemacht, bevor Sie mit dem GH begonnen haben?	64,1	25,6	10,3

	Prozent	
	Ja	**Nein**
Falls ja: Spielen diese Erfahrungen eine zentrale Rolle, dass Sie mit dem GH begonnen haben?	61,5	28,2

Die Items des PAGE II beziehen sich generell auf AgEs ohne Spezifizierung des Kontextes (die Anweisung lautet: „*Bitte lesen Sie die folgenden Aussagen aufmerksam durch und kreuzen Sie bei jeder Aussage an, ob Sie die beschriebenen Phänomene in Ihrem Leben nie, fast nie, selten, gelegentlich oder häufig erlebt haben*"). Er wurde für die Klientel der Beratungsstelle des Instituts

für Grenzgebiete der Psychologie und Psychohygiene (IGPP) entwickelt, bei denen die AgEs mit einem Leidensdruck verbunden sind. Für diesen Fragebogen sind zwar keine Standardwerte, jedoch Werte von einigen Vergleichsgruppen verfügbar. In unserem Fall sind zwei von Interesse: die klinische Gruppe des IGPP-Klientels und eine studentische Stichprobe, die in etwa der Normalbevölkerung zu entsprechen scheint (Fach, 2024).[11] Die vier Grundkategorien sind „internale Phänomene", „externale Phänomene", „Koinzidenzphänomene" und „Dissoziationsphänomene" (Belz & Fach, 2015; Fach, 2022, 2024). Wir stellen in den PAGE-II-Daten eine deutliche Dominanz externaler Phänomene fest (siehe Tabelle 10). Dies unterscheidet sie vom typischen Muster einer studentischen Stichprobe, aber auch von der Stichprobe der IGPP-Klientel, bei der das durchschnittliche Maximum bei den Koinzidenzphänomenen und das Minimum bei den Dissoziationsphänomenen zu finden ist. Ihre Durchschnittswerte auf den vier PAGE-II-Subskalen liegen durchweg über den Vergleichswerten der studentischen Stichprobe (Fach et al., 2024, S. 58).

Tabelle 10

Mittelwerte der PAGE-II-Subskalen und der Gesamtskala für Ghosthunter, für die IGPP-Klientenstichprobe und eine studentische Stichprobe

	Gesamt	external	internal	Koinzidenz	Dissoziation
***M* Ghosthunter**	1,44	2,31	1,19	1,61	0,67
(Studentische Stichprobe)	(0,86)	(0,75)	(0,86)	(1,32)	(0,52)
(IGPP-Klientel)	(1,36)	(1,32)	(1,31)	(2,05)	(0,75)
***SD* Ghosthunter**	0,66	0,83	0,82	0,88	0,62
(Studentische Stichprobe)	(0,53)	(0,67)	(0,69)	(0,76)	(0,60)
(IGPP-Klientel)	(0,82)	(1,15)	(1,11)	(1,08)	(0,89)

Anmerkung: Fünf-Punkte-Skala von „nie" (= 0), „fast nie" (= 1), „selten" (= 2), „gelegentlich" (= 3) bis „oft" (= 4). Vergleichswerte einer studentischen Stichprobe ($N = 450$) und einer Stichprobe der IGPP-Klientel ($N = 193$) in Klammern (Fach et al., 2024, S. 58).

11 Es liegen Daten aus einer repräsentativen Schweizer Stichprobe vor, die mit der ersten Version des PAGE erhoben wurden. Die Werte einer mit dem PAGE II erhobenen Studenten-Stichprobe weisen vergleichbare Merkmale auf, sodass davon ausgegangen werden kann, dass sie nahe an den Normwerten liegen (Fach, 2024).

Der Glaube an das Übernatürliche oder Paranormale

Der Glaube an das Übernatürliche bzw. an paranormale Phänomene, gemessen mit der *Belief in the Supernatural Scale* (BitSS; Schofield et al., 2018), ist bei den Ghosthuntern höher als bei einer britischen Validierungsstichprobe der BitSS (N = 700).[12] Dies gilt für alle Subskalen mit Ausnahme der Subskala „Religiöser Glaube" (siehe Tabelle 11).

Tabelle 11

Mittelwerte der BitSS-Subskalen und der Gesamtskala für Ghosthunter sowie einer britischen Validierungsstichprobe

	Gesamt	Mentale und psychische Phänomene	Religiöser Glaube	Psychokinese	Übernatürliche Wesen	Häufige paranormale Wahrnehmungen
M **Ghosthunter**	3,90	5,02	2,43	3,48	3,73	3,83
(Validierungsstichprobe)	(3,24)	(3,80)	(2,72)	(2,56)	(3,38)	(2,96)
SD **Ghosthunter**	1,21	1,46	1,29	1,36	1,46	1,41
(Validierungsstichprobe)	(1,41)	(1,64)	(1,64)	(1,57)	(1,77)	(1,51)

Anmerkung: Sieben-Punkte-Skala von „stimme überhaupt nicht zu" (= 1) bis „stimme voll und ganz zu" (= 7). Vergleichswerte der britischen Validierungsstichprobe (N = 700) in Klammern.

Explorative Korrelationsstatistiken und Statistiken zu Gruppenunterschieden

Soziodemografische Daten und religiöse Orientierung: Wir untersuchten die Daten explorativ auf Korrelationen und Gruppenunterschiede, die Anhaltspunkte für weitere Forschungshypothesen liefern können. Unser besonderes Interesse galt dem Zusammenhang zwischen soziodemografischen Daten, Religionszugehörigkeit, paranormalen oder übernatürlichen Überzeugungen, AgEs und der Faszination für GH sowie den Erklärungsmodellen für ungewöhnliche Phänomene im Kontext von GH.

Die soziodemografischen Daten erwiesen sich als nicht aussagekräftig für die Unterscheidung von Untergruppen, und es wurden keine signifikanten Korrelationen zwischen den oben

12 Malcolm Schofield stellte die Daten aus den Stichproben zweier Validierungsstudien zur Verfügung (Schofield, E-Mail vom 1. Februar 2022). Die Werte beider Studien (N = 382, N = 312) wurden hier zusammengefasst und gemittelt.

aufgeführten Variablen gefunden, was allerdings teilweise auf die geringe Stichprobengröße zurückzuführen sein könnte.

Wir konnten auch keine signifikanten Gruppenunterschiede hinsichtlich der religiösen oder spirituellen Orientierung in den Mittelwerten der BitSS-Skalen feststellen, mit einer Ausnahme: Diejenigen, die sich als Christen identifizieren, wiesen signifikant höhere Werte auf der BitSS-Subskala „religiöser Glaube" auf, die sich auf traditionelle religiöse Überzeugungen bezieht (H [1, N = 39] = 5,034, p = 0,025, η^2= 0,08).

AgEs vor Beginn des GH: Es wurden signifikante Gruppenunterschiede hinsichtlich AgEs vor dem Beginn mit dem GH als Hobby festgestellt. Wir fanden einen hochsignifikanten Gruppenunterschied zwischen früheren AgEs und der Faszination für den Kontakt mit dem Paranormalen oder Übernatürlichen durch GH (H [2, N = 39] = 17,973, p < 0,001, η^2 = 0,44). Um ein klareres Bild zu erhalten, schlossen wir diejenigen aus, die sich nicht sicher waren, ob sie vor dem Beginn mit dem GH AgEs gehabt hatten (siehe Tabelle 9). Die verbleibenden beiden Vergleichsgruppen umfassten Personen, die sicher waren, dass sie vor der Aufnahme von GH paranormale Erfahrungen gemacht hatten (n = 25), und solche, die sicher waren, dass sie keine solchen Erfahrungen gemacht hatten (n = 4). Der Gruppenunterschied hinsichtlich der oben erwähnten Faszination ist signifikant: (H [1, N = 29] = 12,963, p < 0,001, η^2 = 0,44) (siehe Tabelle 12).

Tabelle 12

Medianwerte für Skepsis und die Akzeptanz von Erklärungsmodellen bei Ghosthuntern mit und ohne AgEs vor Beginn mit dem GH

Skepsis	Ghosthunter mit AgEs vor GH	Ghosthunter ohne AgEs vor GH
Selbst eingeschätzte Skepsis gegenüber paranormalen Phänomenen	60,00	86,50
Erklärung		
Akzeptanz einer „spiritualistischen" Erklärung	4,0	3,0
Akzeptanz der Erklärung „Location"	4,0	3,0
Erklärung zur Akzeptanz „Team"	4,0	3,0

Anmerkung: Die Skepsis wurde auf einer Skala von 0 („überhaupt nicht") bis 100 („äußerst skeptisch") gemessen. Die Akzeptanz der Erklärungsmodelle wurde mit einer Fünf-Punkte-Skala gemessen (1 = „Ich stimme überhaupt nicht zu", 5 = „Ich stimme voll und ganz zu").

Wir verglichen diese beiden Gruppen hinsichtlich ihrer selbst eingeschätzten Skepsis gegenüber paranormalen Phänomenen. Der Unterschied liegt nahe am Signifikanzniveau: H [1, N = 29] = 3,753, p = 0,053, η^2 = 0,10.

Der Unterschied zwischen diesen beiden Gruppen bei der Akzeptanz des „spiritualistischen" Erklärungsmodells für paranormale Phänomene bei PUs beträgt H [1, N = 29] = 7,872, p = 0,005, η^2 = 0,25; der Unterschied in der Akzeptanz der „Standort"-Erklärung beträgt H [1, N = 29] = 10,947, p < 0,001, η^2 = 0,37. Bei der dritten Erklärung, der „Team"-Erklärung, fanden wir keinen signifikanten Unterschied. Die Medianwerte unterscheiden sich jedoch.

PAGE II: Wir fanden zwei signifikante Korrelationen zwischen zwei PAGE-II-Subskalen und der Zustimmung zur „spiritualistischen" Erklärung sowie zur „Orts"-Erklärung (siehe Tabelle 13), eine signifikante Korrelation zwischen der PAGE-II-Subskala „Dissoziationsphänomene" und der Zustimmung zur „Orts"-Erklärung (r = 0,396; p = 0,013) sowie eine weitere signifikante Korrelation zwischen der PAGE-II-Subskala „Zufallsphänomene" und der Zustimmung zur „spiritualistischen" Erklärung (r = 0,393; p = 0,013).

Tabelle 13

Korrelationen zwischen der Zustimmung zu Erklärungsmodellen oder -theorien und den PAGE-II-Subskalen sowie der Gesamtskala

Erklärung	Gesamt	external	internal	Koinzidenz	Dissoziation
Ich sehe sie als Zeichen von Verstorbenen, Geistern oder anderen Wesenheiten („spiritistisch")	.286	.293	.066	.393*	.244
Sie hängen mit den Mitgliedern des Teams zusammen („Team")	.079	.044	.095	.097	.080
Sie hängen mit der Geschichte der Location zusammen („Location")	.231	.018	.161	.186	.396*

Anmerkung: Spearman-Korrelationskoeffizient; * $p \leq 0{,}05$

BitSS: Der Grad der Übereinstimmung mit der „spiritualistischen" Erklärung korreliert signifikant mit der BitSS-Gesamtskala und vier Subskalen; der Grad der Übereinstimmung mit der „Location"-Erklärung korreliert mit der BitSS-Gesamtskala und zwei Subskalen (siehe Tabelle 14).

Tabelle 14

Korrelationen zwischen der Zustimmung zu Erklärungsmodellen oder -theorien und den BitSS-Subskalen sowie der Gesamtskala

Erklärung	Gesamt	Mentale und psychische Phänomene	Religiöser Glaube	Psychokinese	Übernatürliche Wesen	Häufige paranormale Wahrnehmungen
Ich sehe sie als Zeichen von Verstorbenen, Geistern oder anderen Wesenheiten („spiritistisch")	.547***	.548***	.345*	.108	.433**	.552***
Sie hängen mit den Mitgliedern des Teams zusammen („Team")	.007	.084	-.199	.109	-.192	.081
Sie hängen mit der Geschichte der Location zusammen („Location")	.357*	.382*	.028	.126	.414**	.292

Anmerkung: Spearman's *r*; * $p \leq 0{,}05$, ** $p \leq 0{,}01$, *** $p \leq 0{,}001$.

Diskussion 2

GH ist ein kostspieliges Hobby. Eine überdurchschnittliche Beschäftigungsquote von etwa 90 % im Vergleich zu 68 % in der Allgemeinbevölkerung (im Alter zwischen 25 und 74 Jahren) ist daher nicht überraschend. Sie liegt auch über dem Durchschnitt der PAGE-Stichprobe der IGPP-Klientel[13] mit 49 % Beschäftigungsquote (Fach, 2022). Der Anteil der Personen in einer festen Beziehung ist unter Ghosthuntern höher als in der Allgemeinbevölkerung: 72 % der Teilnehmenden haben einen festen Partner. Der Durchschnittswert für die Allgemeinbevölkerung liegt bei etwa 50 %. (siehe FN 7). Dies könnte als Hinweis darauf interpretiert werden, dass das Hobby Ghosthunting für Menschen in prekären sozialen Situationen nicht besonders gut geeignet ist.

Der Begriff „Ghosthunter" impliziert den Glauben an die Existenz von Geistern verstorbener Personen oder anderer „übernatürlicher" Wesen, die „gejagt" werden können. Bei genauerer Betrachtung erweist sich dies jedoch als nicht ganz so einfach. Die Bezeichnung für diese Freizeitbeschäftigung wurde aus dem englischsprachigen Raum unübersetzt übernommen,

13 Die „IGPP-Klientel" ist die Klientel, die sich an die Beratungsstelle des Instituts für Grenzgebiete der Psychologie und psychischen Gesundheit (IGPP) gewandt hat.

ohne dass dies zwangsläufig auch für die damit verbundenen Glaubensvorstellungen gilt. Der Anteil der Bevölkerung, der an „übernatürliche geistige Wesenheiten (z. B. Engel, Dämonen, Feen, Geister usw.)“ glaubt, unterscheidet sich deutlich zwischen den Vereinigten Staaten (61 %) und Deutschland (33 %) (Ipsos Global Advisor, 2023).[14] Insgesamt scheint die deutsche GH-Szene den säkularisierten Zustand der deutschen Gesellschaft widerzuspiegeln. Der Anteil der Christen (36 %) ist geringer als in der Gesamtbevölkerung, der laut dem „Religionsmonitor“ bei 50 % liegt (Müke et al., 2023, S. 16). Überraschend ist, dass 21 % angaben, an nichts Göttliches zu glauben. Offenbar schließen GH und Atheismus einander nicht zwangsläufig aus. Auch der Anteil der „sonstigen“ Religionszugehörigkeit ist überraschend hoch: 21 % der Teilnehmenden kreuzten diese Option im Fragebogen an. Dies könnte bedeuten, dass ein großer Teil dieser Menschen sich keiner Religionsgemeinschaft zugehörig fühlt und nicht besonders spirituell orientiert ist, aber dennoch an ein göttliches Prinzip glaubt. Zählt man die Werte „Atheist“, „Agnostiker“ und „Sonstige“ zusammen, ergibt sich ein Anteil von 50 %. Der kombinierte Anteil dieser Gruppe in der Gesamtbevölkerung beträgt 40 % (Müke et al., 2023, S. 16).[15] Der Anteil von Menschen, die islamischen Religionsgemeinschaften angehören, beträgt in der deutschen Gesamtbevölkerung 8,5 %. In unserer Stichprobe waren sie nicht vertreten.

Man könnte meinen, dass Ghosthunter stärker als der Durchschnitt an die Existenz von Geistern, Dämonen und anderen übernatürlichen Wesen glauben und daher tendenziell eine religiösere Weltanschauung haben, doch scheint der Zusammenhang nicht so einfach zu sein, wie aus dem Vergleich der Werte der BitSS-Subskalen „Religiöser Glaube“ und „Übernatürliche Wesen“ der Ghosthunter-Stichprobe und der britischen Validierungsstichprobe hervorgeht (siehe Tabelle 11). Besonders auffällig ist die Subskala „Religiöser Glaube“. Hier weist die britische Stichprobe einen höheren Mittelwert auf, obwohl laut einer Ipsos-Umfrage aus dem Jahr

14 Diese Zahlen stammen aus einer Online-Umfrage mit 19.731 Befragten aus 26 Ländern. Sie sind nicht repräsentativ. Andere Zahlen aus repräsentativen Umfragen in den USA und Deutschland sind nicht direkt vergleichbar, da die Erhebungszeiträume unterschiedlich sind und die Fragen nicht identisch waren. Das Gesamtbild bleibt jedoch konsistent. Laut einer repräsentativen Umfrage von Ipsos aus dem Jahr 2019 glauben 46 % der Amerikaner, dass es Geister gibt (Ipsos, 2019), während dies laut einer repräsentativen Umfrage aus dem Jahr 2002 nur auf 8 % der deutschen Bevölkerung zutraf (Institut für Demoskopie Allensbach, 2002). Die Zahl für Deutschland mag inzwischen gestiegen sein, liegt aber sicherlich deutlich unter der für die USA.

15 Die Werte sind jedoch nicht genau vergleichbar und sollten nur als Richtwerte verstanden werden, da in der Bevölkerungsbefragung des „Religionsmonitors“ keine Mehrfachantworten möglich waren, wie es bei uns der Fall war. Zudem enthielt der „Religionsmonitor“ nicht die Kategorien „ungebundene Spiritualität/Religiosität“, „Atheist“ und „Agnostiker“ als Antwortmöglichkeiten, sondern stattdessen „andere Religionsgemeinschaft“ und „keine Religionsgemeinschaft“ (Müke et al., 2023, S. 16).

2023 die Religiosität im Vereinigten Königreich sogar etwas geringer ist als in Deutschland (Ipsos Global Advisor, 2023).

Der größte Unterschied zwischen den beiden Stichproben besteht in der Subskala „Mentale und psychische Phänomene", doch ist der Unterschied auch in den beiden Subskalen „Psychokinese" und „Allgemeine paranormale Wahrnehmungen" beträchtlich (siehe Tabelle 11). Diese drei Subskalen beziehen sich auf Phänomene, die während der GH erlebt werden können, wobei der Faktor „Allgemeine paranormale Wahrnehmung" sowohl Items zum Thema Spuk als auch eine Reihe von Items zur Wahrsagerei (Astrologie, Kartenlegen) umfasst. Wir können daher mit einiger Sicherheit sagen, dass Ghosthunter stärker als der Durchschnitt der Bevölkerung von paranormalen oder übernatürlichen Phänomenen überzeugt sind. Dies ist zu erwarten, da das GH darauf ausgerichtet ist, paranormale Phänomene zu erleben.

Es ist bekannt, dass die Vorstellung, die individuellen Glaubensvorstellungen und Weltanschauungen von Menschen müssten konsistent und frei von Widersprüchen sein, falsch ist (z. B. Luhrmann, 1989). Dies spiegelt sich auch in den Erklärungen der Ghosthunter für paranormale Phänomene wider. Etwa 85 % sehen in ihnen Zeichen verstorbener Personen oder übernatürlicher Wesen – dies trotz des relativ hohen Anteils an Atheisten und Agnostikern unter ihnen. Wie zu erwarten, wird ein starker Zusammenhang zwischen den erlebten Phänomenen und der Geschichte der Locations angenommen, da dies dem grundlegenden Narrativ des GH entspricht. Ghosthunter besuchen bestimmte Orte, weil sie glauben, dass solche Phänomene dort wahrscheinlicher als anderswo auftreten, oder weil andere Menschen dort AgEs hatten, die sie mit der Location in Verbindung bringen. Die Mehrheit geht zumindest von einer Art „Ortsgedächtnis" oder „lebendigen Geschichte eines Ortes" aus. Dies steht im Einklang mit dem hohen Anteil derjenigen, die von der Geschichte der PU-Locations fasziniert sind (siehe Tabelle 7). Aus psychodynamischer Perspektive lassen sich viele paranormale Phänomene als Externalisierung psychologischer oder emotionaler Spannungen interpretieren (Belz & Fach, 2015). Mehr als die Hälfte der Teilnehmenden (53,9 %) stimmt dieser Erklärung ebenfalls teilweise oder voll zu („Team"). Trotz dieses starken Glaubens an das Paranormale steht die Mehrheit der Menschen den von ihnen erlebten Phänomenen skeptisch gegenüber. Dies kann als Ausdruck eines wissenschaftlichen Selbstverständnisses interpretiert werden. Es steht im Einklang mit früheren Ergebnissen aus Interviews mit deutschen Ghosthuntern (Bartoschek & Waschkau, 2013, S. 38–40).

AgEs spielen für viele eine bedeutsame Rolle für die Entscheidung, dem GH als Freizeitbeschäftigung nachzugehen. Ghosthunter haben es sich zum Hobby gemacht, an ausgewählten Locations AgEs zu erleben, die in erster Linie als extern wahrgenommen werden. Es überrascht daher kaum, dass wir in den PAGE-II-Daten eine deutliche Dominanz externaler Phänomene

feststellen (siehe Tabelle 10). Dies unterscheidet die Ghosthunter von anderen Stichproben, bei denen der PAGE II angewendet wurde (Fach, 2024). Ein weiterer Unterschied zwischen der IGPP-Klientel (siehe FN 13) und den Ghosthuntern liegt im Bereich der sozialen Bindungen (soziale Situation, Beziehung). 45 % der IGPP-Klientel leben ohne feste Partnerschaft (Fach, 2024, S. 354), während 72 % der Teilnehmenden unserer Studie einen festen Lebenspartner haben. Wolfgang Fach hat in seinen Untersuchungen einen signifikanten Zusammenhang zwischen der Häufigkeit des Erlebens von AgEs und der Qualität der Bindung in Beziehungen, der sozialen Integration und auch der Beschäftigungssituation der Betroffenen festgestellt. Die Verallgemeinerbarkeit seiner Befunde wird durch die kleine und sehr spezielle Stichprobe der Ghosthunter etwas relativiert. Bei Ghosthuntern ist das Erleben von AgEs offensichtlich nicht mit Leidensdruck oder negativen Gefühlen verbunden, sondern weckt bei vielen den Wunsch, solche Erfahrungen zu vertiefen und aktiv zu suchen. Schließlich gaben fast zwei Drittel unserer Teilnehmenden an, dass frühere AgEs eine entscheidende Rolle dabei gespielt hätten, dass sie sich dem GH als Freizeitbeschäftigung zugewandt hätten.

Es liegt daher auf der Hand, dass für fast alle Ghosthunter der mögliche Kontakt mit dem Paranormalen oder Übernatürlichen eine besondere Faszination an ihrem Hobby darstellt. Angesichts der Bedeutung von vorgängigen AgEs vor dem Beginn mit dem GH könnte eine genauere Untersuchung der vier Teilnehmenden (10,3 %) vielversprechend sein, die vor dem Einstieg in das GH keine paranormalen Phänomene erlebt haben. Ein statistischer Gruppenvergleich, bei dem eine Gruppe so klein ist, ist natürlich problematisch. Alle daraus gewonnenen Erkenntnisse sind daher höchst spekulativ. Dennoch haben wir einige Gruppenvergleiche durchgeführt, und das daraus resultierende Bild ist relativ konsistent. Die interessantesten statistischen Ergebnisse bezogen sich auf die Gründe für die Faszination an GH, auf die Theorien oder Erklärungen für die paranormalen Phänomene während der PUs sowie auf paranormale oder übernatürliche Überzeugungen (BitSS-Skalen). Wir fanden einen hochsignifikanten Gruppenunterschied zwischen früheren paranormalen Erfahrungen und der Faszination durch den Kontakt mit dem Paranormalen oder Übernatürlichen mittels GH. Bei den vier Personen ohne AgEs vor dem GH machen etwaige paranormale Phänomene, die während der PUs auftreten können, einen deutlich geringeren Anteil an der Faszination aus als bei der überwiegenden Mehrheit der Ghosthunter. Es liegt die Vermutung nahe, dass hier ein Zusammenhang mit Skeptizismus besteht, und tatsächlich weisen diese vier einen höheren Durchschnittswert auf, wenn es darum geht, ihre eigene Skepsis gegenüber paranormalen Phänomenen einzuschätzen (siehe Tabelle 12).

Ein ähnliches Bild zeigt sich bei den Erklärungsmodellen für paranormale Phänomene während der PUs. Die hochsignifikanten Korrelationen zwischen dem Grad der Übereinstim-

mung mit zwei der drei angebotenen Erklärungsmodelle und paranormalen oder übernatürlichen Überzeugungen sind nicht überraschend, da sie mit grundlegenden Vorstellungen von GH zusammenhängen. Allerdings kamen sowohl die „spiritistische" Erklärung als auch die „Location"-Erklärung für die vier Personen, die keine vorgängige paranormalen Erfahrungen gemacht hatten, praktisch nicht in Frage. Der Unterschied zwischen den beiden Gruppen mit und ohne AgEs vor dem GH verschwindet im dritten, dem „Team"-Erklärungsmodell, wonach die Phänomene mit den Mitgliedern des Teams zusammenhängen, was neben der animistisch-psychologischen Erklärung auch als konventionelle Erklärung interpretiert werden kann, nämlich dass es sich um Täuschungen oder Selbsttäuschungen handeln könnte. Es scheint, dass selbst Personen, die im Allgemeinen skeptisch gegenüber der Existenz paranormaler Phänomene eingestellt sind, Freude am GH haben können (siehe Tabelle 12).

Was die Ergebnisse der Statistik der BitSS-Skalen und die Übereinstimmung mit den drei Erklärungsmodellen betrifft, sind die nicht signifikanten Korrelationen interessanter als die erwarteten signifikanten (siehe Tabelle 13). Es bestehen keine signifikanten Korrelationen zwischen der Zustimmung zur „spiritistischen" Erklärung und zur „Location"-Erklärung und dem Glauben an psychokinetische Phänomene. Dies ist interessant, da im GH auch sogenannte Bewegungstests eingesetzt werden, bei denen sich Objekte ohne sichtbare Ursache bewegen sollen. Wird dies tatsächlich beobachtet, könnte es aus einer externen Perspektive als psychokinetisches Phänomen interpretiert werden, d. h. als Ergebnis der Wechselwirkung eines menschlichen Bewusstseins (Willensanstrengung) mit einem externen physischen Objekt. Diese Sichtweise passt jedoch nicht zum Grundnarrativ des GH. Obwohl der Begriff „paranormale Aktivität" oft neutral verwendet wird, wird allgemein angenommen, dass externe Akteure die Ursache sind und nicht die Ghosthunter selbst, wie es die Fragen der BitSS-Skala „Psychokinese" nahelegen würden. Dies spiegelt sich auch in einer deutlich geringeren Zustimmung zu der Erklärung wider, dass die paranormalen Phänomene während des GH mit Mitgliedern des Teams in Zusammenhang stehen könnten. Kurz gesagt: Wenn sich Objekte von selbst bewegen, dann sind es die Geister und nicht PK.[16]

Die beiden signifikanten Korrelationen zwischen zwei PAGE-II-Subskalen und der Zustimmung zur „spiritistischen" Erklärung sowie zur „Location"-Erklärung sind schwer zu interpretieren (siehe Tabelle 14). Wir fanden eine signifikante Korrelation zwischen der PAGE-II-Subskala „Dissoziationsphänomene" und der Zustimmung zur Erklärung „Location" sowie

16 In einer deutschen Bevölkerungsbefragung aus dem Jahr 2025, in der unter anderem nach der Vorstellbarkeit von Psi-Phänomenen gefragt wurde, erzielte die Psychokinese mit 17 % die niedrigste Zustimmungsrate. Die übersinnliche Wahrnehmung im Zusammenhang mit Tod und Krisen erhielt mit 71 % Zustimmung den höchsten Wert (Schmied-Knittel, im Druck).

eine weitere signifikante Korrelation zwischen der PAGE-II-Subskala „Koinzidenzphänomene" und der Zustimmung zur „spiritistischen" Erklärung. Ein Hauptproblem bei der Interpretation besteht darin, dass der PAGE II Erfahrungen misst, die Erklärungsmodelle jedoch Überzeugungen oder theoretische Annahmen über die Welt betreffen.[17] Ein weiterer wichtiger Punkt kann leicht übersehen werden, wenn man nur die signifikanten Korrelationen betrachtet. Die größte mittlere Differenz zwischen der GH-Stichprobe und den Vergleichsstichproben besteht bei der PAGE-II-Subskala „externale Phänomene". Dies war aufgrund des Ziels des GH zu erwarten. Es besteht jedoch nur eine schwache Korrelation mit der spiritistischen Hypothese, die das gewählte Signifikanzniveau von 5 % nicht ganz erreicht (r = .293; p = .071). Bedenkt man, dass Ghosthunter versuchen, das Unsichtbare mit verschiedenen technischen Hilfsmitteln sichtbar zu machen und sinnvolle Verbindungen zwischen spirituellen und materiellen Phänomenen herzustellen, so wird eine signifikante Korrelation zwischen der Häufigkeit des Erlebens von „Koinzidenzphänomenen" und der Zustimmung zu einem spiritistischen Erklärungsmodell verständlich.[18] Für die gefundene signifikante Korrelation zwischen der Häufigkeit des Erlebens von „Dissoziationsphänomenen" und der Zustimmung zur „Location"-Erklärung konnten wir keine solche nachvollziehbare Erklärung finden.

Schlussfolgerung

Da es unseres Wissens keine vergleichbaren Studien gab, mit denen die Daten verglichen werden konnten und die datenbasierte Hypothesen ermöglicht hätten, war diese Studie explorativ. Die Auswertung war aufgrund der geringen Stichprobengröße deskriptiv. Einige der Ergebnisse waren zu erwarten, wie etwa, dass die aktiv nach AgEs suchenden Ghosthunter im Durchschnitt mehr AgEs zu erleben scheinen als eine studentische Vergleichsstichprobe. Der Glaube an das Paranormale oder Übernatürliche war ebenfalls stärker ausgeprägt als bei der Validierungsstichprobe des BitSS. Andere Ergebnisse sind eher unerwartet, wie etwa die Tatsache, dass 85 % an einen Zusammenhang zwischen den vermuteten paranormalen Phänomenen und Verstorbenen oder Geistern glauben, während sich gleichzeitig 30 % als Atheisten oder Agnos-

17 Obwohl man von einer starken Wechselwirkung zwischen Glaubensvorstellungen und Erfahrungen ausgehen muss – Wahrnehmung und damit auch Erfahrungen werden maßgeblich von Glaubensvorstellungen und kulturell vermittelten Konzepten beeinflusst, und Erfahrungen können wiederum Überzeugungen schaffen oder verstärken –, sind Ausmaß und Richtung selten eindeutig identifizierbar und dürften von Individuum zu Individuum stark variieren.

18 „Coincidence phenomena are connections between ordinarily disconnected elements of the self-model and the world-model and often experienced as meaningful links between mental and material events" (Belz & Fach, 2015, S. 371).

tiker bezeichnen. Dies spricht für eine hohe Akzeptanz der Überlebenshypothese, die offenbar nicht an einen Glauben an einen Gott oder das Göttliche gebunden ist. GH, den Glauben an die Seelen Verstorbener implizierend, und Atheismus schließen sich dementsprechend nicht unbedingt aus.

Die Faszination für das Unheimliche, Okkulte, Paranormale und Übernatürliche spiegelt sich in Mythen, Legenden, Werken der Literatur und Filmkunst wider, aber auch in „Geistertouren" als touristischen Angeboten (z. B. Houran et al., 2020; Pedreño-Peñalver et al., 2023), im Internet und in sozialen Netzwerken. Eine neuere Variante zeigt sich im „Virtual Dark Tourism" auf YouTube, d. h. der virtuellen Teilnahme an GH-Aktivitäten als eine Form des „armchair travel [Sesselreisens]" (Basaraba, 2024). Der angenehme Schrecken, der Nervenkitzel oder die Angstlust können in der sicheren Umgebung des Wohnzimmers genossen werden. Dies ist ein Aspekt, der durch GH an vermeintlichen Spukorten noch verstärkt werden kann. Praktisch alle nennen den Kontakt mit dem Paranormalen oder Übernatürlichen als einen besonders faszinierenden Aspekt des GH. Hinzu kommt der Faktor der Authentizität, der bei der medienvermittelten Rezeption des Paranormalen fehlt.[19] Ob darüber hinaus der Impuls, das Paranormale in einem mehr oder weniger wissenschaftlichen Sinne zu erforschen, eine zentrale Motivation ist oder aber der Wunsch, das eigene Glaubenssystem bestätigt zu finden oder es in Frage zu stellen, mag von Person zu Person variieren.

Danksagung

Wir möchten uns bei allen Teilnehmenden der Studie sowie bei den Gutachtern bedanken, die uns sehr wertvolle Anregungen zur Verbesserung des Artikels gegeben haben.

Literatur

Auerbach, L. (2004). *Ghost hunting: How to investigate the paranormal*. Ronin.

Bagans, Z. (2020). *Ghost-hunting for dummies*. John Wiley & Sons, Inc.

Bartoschek, S. J., & Waschkau, A. (2013). *Ghosthunting: Spurensuche im Jenseits*. Alibri.

Basaraba, N. (2024). The rise of paranormal investigations as virtual dark tourism on YouTube. *Journal of*

19 Dies gilt sogar für den „paranormalen Tourismus". Houran et al. schreiben: „Overall, we speculate that paranormal-like experiences are appealing to many people, in part, because they are unique and elusive" (Houran et al., 2020, S. 17).

Heritage Tourism, *19*(2), 287–309. https://doi.org/10.1080/1743873X.2023.2268746

Belz, M., & Fach, W. (2015). Exceptional experiences (ExE) in clinical psychology. In E. Cardeña, J. Palmer, & D. Marcusson-Clavertz (Hrsg.), *Parapsychology: A handbook for the 21st century* (S. 364–379). McFarland & Company, Inc.

Brown, A. (2006). *Ghost Hunters of the South*. University Press of Mississippi.

Fach, W. (2018). *Fragebogen zur Erfassung der Phänomenologie außergewöhnlicher Erfahrungen PAGE-R-II: 2. Revidierte und gekürzte Fassung*.

Fach, W. (2022). Exceptional experiences (ExE) and bonding styles: Autonomy and bonding as basic human needs and as structural determinants of ExE. *Psychotherapy Section Review*, *1*(67), 12–41. https://doi.org/10.53841/bpspsr.2022.1.67.12

Fach, W. (2024). *Das Spektrum des Außergewöhnlichen: Konzeptionelle Ansätze, empirisch-phänomenologische Untersuchungen und plananalytische Fallstudien zur mentalen Repräsentation bei außergewöhnlichen Erfahrungen* [Dissertation, Universität Bern]. https://doi.org/10.48549/5179

Fach, W., Alonso, Y., Evrard, R., Simmonds-Moore, C., & Tressoldi, P. E. (2024). Counseling psychology practice and research: International studies with the PExE-II. In J. Kornmeier (Hrsg.), *Biennial report 2022–2023* (S. 57–59). Institut für Grenzgebiete der Psychologie und Psychohygiene.

Hanks, M. (2015). *Haunted heritage: The cultural politics of ghost tourism, populism, and the past*. Left Coast Press, Inc.

Hanks, M. (2016a). Between electricity and spirit: Paranormal investigation and the creation of doubt in England. *American Anthropologist*, *118*(4), 811–823. https://doi.org/10.1111/aman.12684

Hanks, M. (2016b). Redefining rationality: Paranormal investigators' humour in England. *Ethnos: Journal of Anthropology*, *81*(2), 262–289. https://doi.org/10.1080/00141844.2014.956775

Hill, S. A. (2017). *Scientifical Americans: The culture of amateur paranormal researchers*. McFarland & Company, Inc., Publishers.

Houran, J., Hill, S. A., Haynes, E. D., & Bielski, U. A. (2020). Paranormal Tourism: Market Study of a Novel and Interactive Approach to Space Activation and Monetization. *Cornell Hospitality Quarterly*, *61*(3), 287–311. https://doi.org/10.1177/1938965520909094

Hufford, D. J. (1982). *The terror that comes in the night: An experience-centered study of supernatural assault traditions*. University of Pennsylvania Press.

Institut für Demoskopie Allensbach. (2002). *Gespenster und böse Geister. Stehen wir vor einer Zunahme des Glaubens an das Irrationale?* http://www.ifd-allensbach.de/uploads/tx_reportsndocs/prd_0201.pdf

Ipsos. (2019). *Less than half of Americans believe ghosts are real*. Ipsos. https://www.ipsos.com/en-us/news-polls/halloween-2019

Ipsos Global Advisor. (2023). *Global religion 2023: Religious beliefs across the world. A 26-country Global Advisor survey*. Müke.

Luhrmann, T. M. (1989). *Persuasions of the witch's craft. Ritual magic in contemporary England*. Harvard University Press.

Mayer, G. (2010). Die Geisterjäger kommen: Phänomenologie der Ghost Hunting Groups. *Zeitschrift für Anomalistik*, *10*(1+2), 17–48.

Mayer, G. (2013). A sample phenomenology of the ghost hunting scene in the USA and in Germany. In O. Jenzen & S. R. Munt (Hrsg.), *The Ashgate research companion to paranormal cultures* (S. 363–376). Ashgate.

Mayer, G. & Schetsche, M. (2019). RSPK investigations. In G. Mayer (Hrsg.), *N equals 1: Single case studies in anomalistics* (S. 87–123). LIT.

Müke, M., Tranow, U., Schnabel, A., & El-Menouar, Y. (2023). *Zusammenleben in religiöser Vielfalt: Warum Pluralität gestaltet werden muss*. https://doi.org/10.11586/2023026

Nowara, S. & Mayer, G. (2024). Deutsches Ghosthunting – Historischer Überblick und Einblicke einer Insiderin. *Journal of Anomalistics / Zeitschrift für Anomalistik*, *24*(1), 151–188. https://doi.org/10.23793/zfa.2024.151

Pedreño-Peñalver, C., Huertas-Valdivia, I., & Orea-Giner, A. (2023). An exploratory study of the future paranormal tourist experience on ghost tours. *Journal of Tourism Futures*, *10*(3), 524–538. https://doi.org/10.1108/JTF-04-2023-0097

Potts, J. (2004). Ghost hunting in the twenty-first century. In J. Houran (Hrsg.), *From shaman to scientist: Essays on humanity's search for spirits* (S. 211–232). The Scarecrow Press.

Schofield, M. B., Baker, I. S., Staples, P., & Sheffield, D. (2018). The Creation and Validation of the Belief in the Supernatural Scale. *Journal of Parapsychology*, *82*(1), 41–64.

Tobacyk, J. J. (2004). A Revised Paranormal Belief Scale. *International Journal of Transpersonal Studies*, *23*(1), 94–98. https://doi.org/10.24972/ijts.2004.23.1.94

Voss, E. (2021). Hospitality and proof: Human mediums, technical media, and controversial knowledge in ghost-hunting practices in the United States. In D. Espírito Santo & J. Hunter (Hrsg.), *Mattering the invisible: Technologies, bodies, and the realm of the spectral* (S. 132–152). Berghahn. https://www.academia.edu/77814708/Hospitality_and_Proof_Human_Mediums_Technical_Media_and_Controversial_Knowledge_in_Ghost_Hunting_Practices_in_the_United_States

Wilson, V. (2005). *Ghost tech: The essential guide to paranormal investigation equipment*. Whitechapel.

English Abstract

Ghost Hunting Groups and Ghost Hunters in Germany: Results of Two Online Surveys

Abstract – We conducted two online surveys with German ghost hunting groups (GHGs) and individual ghost hunters. The first questionnaire targeted entire teams, the second individual members. Our goals were (1) to understand the current GHG scene in Germany – their composition, investigation approaches, and exceptional experiences (ExEs) – and (2) to gain insight into the people engaged in this time- and resource-intensive hobby. We developed two questionnaires and included two established scales in the individual version to assess exceptional experiences (Fragebogen zur Phänomenologie außergewöhnlicher Erfahrungen [Questionnaire on the Phenomenology of Exceptional Experiences, PExE II]) and paranormal beliefs (Belief in the Supernatural Scale [BitSS]). We received twenty-one team responses and thirty-nine from individuals. Results from study 1 showed that GHGs exhibited similarities in site selection, investigative methods, use of technical equipment, and types of recorded anomalies. Social media now largely replaces traditional websites and local TV features. However, teams differ in documentation practices and how they handle collected data. Group sizes, founding dates, and annual investigation frequency vary widely. Study 2 found that ghost hunters reported more ExEs than student samples and hold stronger paranormal beliefs (BitSS). Surprisingly, 85% believe in a link between paranormal phenomena and the deceased, while 30% identify as atheists or agnostics – suggesting belief in spirits of the deceased is not necessarily tied to belief in God. For ghost hunters, ExEs are typically not distressing but inspire curiosity and further pursuit. Nearly two-thirds stated that prior ExEs were key motivators for taking up ghost hunting. In addition to the entertainment aspect, the possibility of 'real' contact with the supernatural plays an important role for many ghost hunters, something that cannot be found in media offerings on this topic. The motivation of ghost hunters ranges from scientific interest to the desire to confirm or question their own beliefs.

Keywords: ghost hunting groups, extraordinary experiences, exceptional experiences, paranormal beliefs, technical equipment, paranormal investigations

Zeitschrift für Anomalistik
Journal of Anomalistics

www.anomalistik.de www.igpp.de

Kommentar zu

Sybo Schouten: *The Use of Psychics in Police Investigations of Missing Persons*

In: *Zeitschrift für Anomalistik*, 25(2) (2025), 306–389

Mehr Skepsis als Zustimmung – deutsche Landeskriminalämter zum Einsatz von Hellsehern bei der kriminalistischen Arbeit

Ulrich Magin[a]

(a) Unabhängiger Autor
Magin.Ulrich@web.de

Zu dem Beitrag von Sybo Schouten, „The Use of Psychics in Police Investigations of Missing Persons" in der *Zeitschrift für Anomalistik 25*(2) möchte ich kurz einige Zeilen beitragen.

Anlässlich einer journalistischen Recherche für eine Publikumszeitschrift wollte ich Anfang 2025 herausfinden, ob auch in Deutschland Hellseher der Kriminalpolizei bei der Aufklärung von Verbrechen helfen. Da der Umfang bei solchen Beiträgen beschränkt ist, beschloss ich, über „Google News" nach Fällen aus jüngster Zeit zu suchen, bei denen Zeitungen Wortmeldungen von Hellsehern bei Mordverdachts- und Vermisstenfällen anführten. Von diesen kamen nur solche Fälle in Betracht, die danach gelöst wurden, um die Aussagen bewerten zu können. So kam ich auf drei Fälle aus dem letzten Jahrzehnt.

Zuerst wurden die Aussagen der Hellseher protokolliert, danach die zuständigen Dienststellen der Kriminalpolizei (entweder per

Telefon oder per E-Mail) befragt, wie die Wortmeldung des Hellsehers eingeschätzt worden war. Das ist Gegenstand meines Artikels und ohne Zweifel keine wissenschaftliche Studie. Die Anfrage kam nur mit journalistischem Anspruch daher, die Ergebnisse sind noch nicht veröffentlicht, werden aber in diesem Jahr erscheinen.

Neben den in den Einzelfällen zuständigen lokalen Polizeidienststellen befragte ich zusätzlich die für das jeweilige Bundesland zuständigen Landeskriminalämter. Auch das war eine journalistische Anfrage, allerdings ergaben sich so Stellungnahmen zu dem Themenkomplex, die sicherlich in dieser Aktualität sonst nicht verfügbar sind. Ich möchte sie hier kurz mitteilen.

Alle betreffenden Landeskriminalämter erhielten Anfang Januar 2025 dieselbe Anfrage, die zum Teil lautete:

„Immer wieder geistern durch die Medien Berichte, nach denen Hellseher der Kriminalpolizei helfen. Ich würde mich freuen, wenn Sie mir dazu kurz – wenn möglich – ein paar Fragen beantworten könnten:

- Existieren Vorschriften, wie mit solchen Angeboten vorgegangen werden soll?
- Wie oft wird ein solches Angebot unterbreitet – gibt es eine Statistik, aus der man eine solche Zahl ersehen könnte?
- „Jedem Hinweis wird nachgegangen" – gilt das auch für Hinweise durch Hellseher?
- Falls solche Hinweise als Indizien statistisch erfasst werden – welchen Prozentsatz bilden die Treffer?

Ich bin mir bewusst, dass dies sehr seltsame Fragen sind und sie vielleicht auch zum größten Teil nicht oder nur schwer beantwortet werden können."

Nur, um dem Bericht von Sybo Schouten eine deutsche Perspektive beizufügen, liste ich im Folgenden die Antworten auf:

Landeskriminalamt Mecklenburg-Vorpommern

Antwort von Kriminalhauptkommissar Niels Borgmann, Pressestelle, 13. Januar 2025:

„Wir haben diese sowohl im eigenen Hause (im LKA MV), als auch in den beiden Polizeipräsidien in MV auf das Vorhandensein von entsprechenden Vorgängen in der Vergangenheit prüfen lassen. Im Ergebnis kann ich Ihnen hiermit mitteilen, dass hier keine Ermittlungsvorgänge für die Landespolizei bekannt sind, bei denen es zu Angeboten von sog. Hellsehern zur Sach-

verhaltserhellung gekommen ist. Somit ist es auch zu keiner entsprechenden Zusammenarbeit gekommen.

Es liegen keine expliziten Vorschriften oder ähnliches zu diesem Bereich bzw. Thema vor."

Landeskriminalamt Nordrhein-Westfalen

Antwort von Maren Menke, Pressesprecherin, 9. Januar 2025

„Bei uns im Landeskriminalamt NRW liegen keine Erkenntnisse vor, dass bei Ermittlungen in NRW mit sogenannten ‚Hellsehern' zusammengearbeitet wurde.

Ich kann Ihnen ausschließlich aus Sicht unserer Pressestelle mitteilen, dass es in den vergangenen Jahren ein/zwei vereinzelte Personen gab, die sich mit einer E-Mail an uns gerichtet haben und ihre Fähigkeiten zur Unterstützung angeboten haben. Diese Angebote zielten aber nicht auf konkrete Ermittlungen ab, sondern waren vielmehr die Bitte um Aufträge gegen entsprechende Bezahlung. Wir haben Absagen erteilt.

Inwiefern auch bei den einzelnen Kreispolizeibehörden solche E-Mails eingegangen sind, kann von hier aus nicht beantwortet werden."

Landeskriminalamt Niedersachsen

Telefonische Auskunft, 8. Januar 2025, Gesprächsnotizen:

„Solche Fälle der Zusammenarbeit sind nicht bekannt. Grundsätzlich wird jedem Hinweis nachgegangen, der nachvollziehbar ist – allerdings nicht Visionen oder Träume, es muss schon faktenbasiert sein.

Ein Hellseher-Hinweis wird als Indiz statistisch nicht erfasst."

Man verwies auf den Beitrag „Psychic Detectives – auch in Deutschland?" von Dr. Michael Schetsche in der Zeitschrift „Die Kriminalpolizei", den die Gewerkschaft der Polizei herausgibt, aus dem Dezember 2007.

Landeskriminalamt Bayern

Telefonische Auskunft, 8. Januar 2025, Gesprächsnotizen:

„Es gibt im LKA keine Statistik zu dem Thema.

Es gibt wohl einige Fälle, bei denen sich Hellseher melden, doch diese werden nicht wirklich ernst genommen, also Fälle, bei denen sich Hellseher oder Wünschelrutengänger melden. Doch diese sind stets kompletter Humbug.

Wenn die Sendung ‚XY ungelöst' im Fernsehen ungelöste Fälle aufgreift, melden sich immer wieder Hellseher."

Abschließend: „Mir ist kein Fall in den letzten Jahren bekannt, in der ein Tipp eines Hellsehers zur Ermittlung des Täters geführt hat."

Es sieht demnach so aus, als habe das Thema des Hellsehers als Helfer der Polizei keine große Relevanz. Dasselbe ergab sich übrigens auch bei den Anfragen zu konkreten Mord- und Vermisstenfällen bei den zuständigen örtlichen Dienststellen.

Zeitschrift für Anomalistik
Journal of Anomalistics

www.anomalistik.de www.igpp.de

Nachruf auf Prof. Dr. phil. Dieter Vaitl (1940–2026)

Jürgen Kornmeier[a]

(a) Institut für Grenzgebiete der Psychologie und Psychohygiene, Freiburg i. Br.

Als Prof. Vaitl 2001 Direktor des IGPP wurde, war ich selbst Doktorand am IGPP, in der Abteilung Theorie und Datenanalyse von PD Dr. Harald Atmanspacher. Damals war das IGPP durch die Zuwendungen der Holler-Stiftung noch sehr gut ausgestattet. Es gab mehrere Abteilungen und dreimal mehr Mitarbeiterinnen und Mitarbeiter als heute. Das Gießener An-Institut, das „Bender Institute of Neuroimaging" nahm gerade seinen Betrieb auf. Dann reduzierten sich jedoch die jährlichen Zuwendungen der Holler-Stiftung plötzlich drastisch, und von nun an war am IGPP Sparen angesagt. Als ich 2008 nach meiner Postdoc-Zeit zurück ans Institut kam, hatte

sich die finanzielle Situation weiter verschärft, was Prof. Vaitl 2013 schließlich zu Umstrukturierungen und zur Ablösung des BION vom IGPP nötigte. Als ich im Juli 2022 die Leitung des IGPP übernahm, war die Zahl der Mitarbeitenden mittlerweile so weit geschrumpft, dass nun am Ende des Jahres wieder eine schwarze Null möglich war. In meiner neuen Funktion als Direktor erlangte ich dann detailliertere Einblicke in die finanzielle Entwicklung des IGPP und den enormen Umbruch von großem finanziellem Überfluss zu ebenso großem Mangel in der 20-jährigen Amtszeit Dieter Vaitls.

Als junger Doktorand schärften mir meine wissenschaftlichen Lehrer ein, dass es karriereschädlich sein könne, wenn man sich öffentlich zum Thema Bewusstsein äußere. Während dieses Thema aber mittlerweile im wissenschaftlichen Diskurs fest verankert ist, stehen die Grenzgebiete, mit Themen wie zum Beispiel außersinnliche Wahrnehmung oder Psychokinese, unverändert im Kreuzfeuer der wissenschaftlichen Kritik. Nach wie vor sind Wissenschaftler und Wissenschaftlerinnen im Bereich der Grenzgebiete konfrontiert mit Ignoranz, Skepsis und Ablehnung und mit potenziell toxischer Wirkung auf individuelle wissenschaftliche Karrieren. Es gab und gibt natürlich auch in anderen Gebieten der Wissenschaft immer wieder Tabuthemen mit all den beschriebenen Implikationen. Für die Wissenschaftlerinnen und Wissenschaftler, die in diesen Bereichen arbeiten, sind die Herausforderungen beachtlich, und es stehen immer wieder Lebensplanungen und Schicksale ganzer Familien auf dem Spiel.

Dieter Vaitl besaß die notwendige Widerstandsfähigkeit, die als Direktor des IGPP notwendig ist. 2012 veröffentlichte er einen Sammelband mit dem Titel *Veränderte Bewusstseinszustände* (Vaitl, 2012). Dieser Band behandelt Themen wie zum Beispiel Halluzinationen, Nahtoderfahrungen, außerkörperliche Erfahrungen, mystische Erfahrungen oder Hypnose. Eine Beschäftigung mit diesen Themen hätte in den 90er Jahren vermutlich das Ende der eigenen wissenschaftlichen Karriere bedeutet.

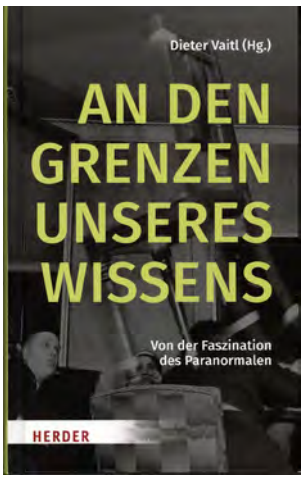
Der 2015 von Prof. Vaitl veröffentlichte Sammelband *An den Grenzen der Erkenntnis* (Mayer et al., 2015) beschäftigt sich mit dem Thema Anomalistik, das die Erforschung allgemein ungewöhnlicher, nicht gut erklärter Anomalien an den Grenzen des Wissens zusammenfasst. Darunter fallen zum Beispiel auch physikalische Anomalien, UFO-Sichtungen oder Astrologie.

Einer von mehreren Gründen für die mangelnde Akzeptanz der Grenzgebiete ist der aktuelle Erkenntnisstand. Es gibt beeindruckende Einzelfallberichte, die aber in der Regel auf Gedächtnisprotokolle von Zeugen beschränkt sind und die sehr oft mehr Fragen aufwerfen, als sie beantworten können. Die Ergebnisse der Laborforschung zeigen Effekte, die zu groß sind, um sie völlig zu ignorieren, aber auch zu klein, um zu überzeugen. Darüber hinaus werden die zuvor gefundenen Effekte bei Replikationen entweder kleiner (der sog. „decline effect"; Protzko & Schooler, 2017) oder verschwinden vollständig (das sogenannte Replikationsproblem, das allerdings nicht auf die Grenzgebiete der Psychologie beschränkt ist (Ioannidis, 2005).

Einen ganz anderen Forschungsansatz verfolgt die Beratungseinheit des IGPP. Hier können Menschen mit belastenden außergewöhnlichen Erfahrungen unentgeltlich Rat einholen. Im Rahmen von Begleitforschung zur Beratungstätigkeit und in Kooperation mit der Abteilung für Psychologie der Universität Freiburg wurde eine Kategorisierung und Quantifizierung der auftretenden Phänomene erarbeitet und dabei interessante und insbesondere auch replizierbare Erkenntnisse gewonnen (Fach, 2024). Dieter Vaitl erkannte schnell das große Potenzial dieses Forschungsansatzes. Als Mitherausgeber der renommierten Buchreihe „Fortschritte der Psychotherapie" konnte er erreichen, dass 2008 ein gesamtes Buch mit dem Titel *Außergewöhnliche Erfahrungen* der Forschung und Theoriebildung der Beratungseinheit des IGPP gewidmet wurde (Belz, 2009).

Es ist dringend notwendig, dass wir Wissenschaftler und Wissenschaftlerinnen unsere Anstrengungen erhöhen, den Menschen unsere Arbeit und die große Bedeutung der Wissenschaft für unsere Gesellschaft verständlich zu erklären. Dass Dieter Vaitl diese wichtige Aufgabe sehr ernst nahm, zeigt sich unter anderem in den immer sehr gut besuchten Vorlesungsreihen, die er über vier Jahre für die Freiburger Bevölkerung organisierte.

Als Prof. Bender 1950 das IGPP gründete, dachte noch niemand an Internet, soziale Medien, künstliche Intelligenz und an die Vielzahl von Möglichkeiten, wie Menschen an Information gelangen können. Niemand konnte sich damals vorstellen, wie einfach es werden würde, zu fotografieren oder nach eigenen Vorstellungen realitätsnahe Bilder zu erzeugen und mit Nachrichten schnell eine große Zahl an Menschen zu erreichen. Prof. Bender verfügte nicht über all diese Möglichkeiten, erkannte aber schon damals die Wirkkraft von Bildern und die Bedeutung der Medien und nutzte die Möglichkeiten seiner Zeit. So engagierte er den Fotografen Leif Geiges, um spektakuläre Spukfälle bildhaft nachzustellen. Diese Bilder, mit hohem künstlerischem Wert, präsentierten Dieter Vaitl und Andreas Fischer 2021 im Kontext des 70-jährigen Jubiläums des IGPP in einer beeindruckenden Ausstellung mit dem Titel „SPUK" im Freiburger Augustinermuseum zusammen mit einem sehr gelungenen gleichnamigen Ausstellungskatalog (Fischer & Vaitl, 2021).



Ebenfalls zum Anlass des Institutsjubiläums und auch zu seinem Abschied als Institutsdirektor gab Dieter Vaitl 2020 den Forschungsband *An den Grenzen unseres Wissens – Von der Faszination des Paranormalen* heraus (Vaitl, 2020). Die ca. 30 Kapitel beschreiben zusammenfassend die interdisziplinäre Institutsarbeit, die in seiner Zeit als Direktor am IGPP entstanden ist.

Von meinen persönlichen Begegnungen mit Prof. Vaitl als Doktorand am Institut wird mir immer seine scherzhafte Standardfrage in Erinnerung bleiben: „Und, Herr Kornmeier, wie sieht es aus? Haben Sie die Weltformel schon entdeckt?" Prof. Vaitl war bis zu seinem Tod Vorsitzender des Vorstands. Für unsere Vorstandssitzungen waren Eberhard Bauer und ich regelmäßig im Hause Vaitl zu Gast. In dieser gemeinsamen Zeit im Vorstand des IGPP durfte ich Dieter Vaitl noch einmal von einer anderen Seite kennen und schätzen lernen. Die Zusammenarbeit mit ihm war immer professionell, vertrauensvoll und von Herzlichkeit und gegenseitiger Wertschätzung geprägt.

Als Institutsdirektor und Vorstandsvorsitzender hat Dieter Vaitl das IGPP über viele Jahre durch herausfordernde Zeiten geführt und entscheidend geprägt. Bis zuletzt lag ihm das IGPP und dessen Zukunft sehr am Herzen.

Der Tod ist natürlich auch ein Grenzgebiet der Psychologie, und die Frage, ob danach noch etwas kommt – das altehrwürdige „survival"-Problem der Society for Psychical Research –, ist auch ein Thema unseres Instituts. Ich würde sagen, es gibt interessante Literatur zu diesem Thema, aber die Frage ist weiterhin offen. Falls der Tod nicht das Ende sein sollte, hat Dieter Vaitl jetzt im Jenseits vielleicht eine Antwort auf die Frage nach der Weltformel. Wir werden ihn jedenfalls im Diesseits vermissen.

Literatur

Belz, M. (2009). *Außergewöhnliche Erfahrunge*n. Hogrefe.

Fach, W. (2024). *Das Spektrum des Außergewöhnlichen: Konzeptionelle Ansätze, empirisch-phänomenologische Untersuchungen und plananalytische Fallstudien zur mentalen Repräsentation bei außergewöhnlichen Erfahrungen* [Bern, Schweiz]. https://boristheses.unibe.ch/5179/

Fischer, A., & Vaitl, D. (Hrsg., zusammen mit dem Institut für Grenzgebiete der Psychologie und Psychohygiene & Augustinermuseum) (2021). *Spuk! Die Fotografien von Leif Geiges*. Michael Imhof.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine, 2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124.

Mayer, G., Schetsche, M., Schmied-Knittel, I., & Vaitl D. (Hrsg.). (2015). *An den Grenzen der Erkenntnis: Handbuch der wissenschaftlichen Anomalistik*. Schattauer.

Protzko, J., & Schooler, J. W. (2017). Decline effects: Types, mechanisms, and personal reflections. In S. O. Lilienfeld & I. D. Waldman (Hrsg.), *Psychological science under scrutiny* (S. 85–107). Wiley. https://doi.org/10.1002/9781119095910.ch6

Vaitl, D. (2012). *Veränderte Bewusstseinszustände*. Schattauer.

Vaitl, D. (Hrsg.).(2020). *An den Grenzen unseres Wissens: Von der Faszination des Paranormalen*. Herder.

Zeitschrift für Anomalistik
Journal of Anomalistics

www.anomalistik.de www.igpp.de

Prof. Dieter Vaitl (1940–2026) am IGPP – eine persönliche Erinnerung

Eberhard Bauer[a]

I

Prof. Dieter Vaitl steht, wenn man so will, in der unmittelbaren Traditionslinie, die noch direkt auf den Institutsgründer Prof. Hans Bender (1907–1991) zurückgeht. Er selbst hat 2001, als er noch Lehrstuhlinhaber für Klinische und Physiologische Psychologie an der Universität Gießen war, dieses Amt von dem frühverstorbenen Professor Johannes Mischo (1930–2001), von 1975 bis 1998 Hans Benders Nachfolger auf dem Lehrstuhl für Psychologie und Grenzgebiete der Psychologie an der Universität Freiburg und – nach Benders Tod 1991 – selbst IGPP-Direktor, übernommen und pendelte mehrere Jahre zwischen Gießen und Freiburg, in der Institutsleitung ‚vor Ort' durch ein Management Board tatkräftig unterstützt.

(a) Institut für Grenzgebiete der Psychologie und Psychohygiene, Freiburg i. Br.

Dieter Vaitls Beziehung zum Institut und zur Freiburger Parapsychologie geht auf jene Jahre zurück, als sich das 1950 gegründete Institut noch auf der „Eichhalde 12" in Freiburg-Herdern auf dem viel beschworenen „Magischen Hügel" befand – er war dort von 1964 bis 1967 als studentische Hilfskraft beschäftigt und hielt diese private Forschungsstätte, ihren Gründer und die ‚anima' des Instituts, die langjährige Geschäftsführerin Frau Lotte Böhringer (1917–1994), immer in bester Erinnerung, wie auch seine Einträge in den Gästebüchern anlässlich seiner Besuche im Eichhalde-Institut zeigen.

Die ersten ‚Arbeitsproben' des stud. phil. Dieter Vaitl in Sachen Grenzgebiete jener Eichhalde-Jahre sind nachzulesen in den Jahrgängen 8 (1965) und 10 (1966) der von Hans Bender 1957 gegrün-

deten *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie* – es handelt sich um zwei Buchbesprechungen sowie um mehrere kommentierende deutsche Übersetzungen der Abstracts von Artikeln aus angloamerikanischen Fachzeitschriften, z. B. dem *International Journal of Parapsychology*, erschienen in der Rubrik „Zeitschriftenspiegel".

Ein weiteres autobiographisches Zeugnis stellt der Brief dar, den er seinem akademischen Lehrer, eben Hans Bender, zu dessen 80. Geburtstag am 5. Februar 1987 geschrieben hat (siehe Abb. 2a–2c); abgedruckt ist der Brief – neben vielen anderen persönlichen Erinnerungen aus Hans Benders Kollegen-, Mitarbeiter- und Freundeskreis – in dem Sonderheft der genannten Zeitschrift (Jg. 29, Heft 1, 1987, S. 90–92), das ich – als Benders ehemaliger Assistent – herausgegeben hatte.

In seinen Erinnerungen an die damalige „Eichhalde-Zeit" figuriert – neben der Teilnahme an den sog. Tonbandeinspielungen mit Konstantin Raudive in Bad Krozingen – auch Dieter Vaitls Begegnung mit dem ‚Spukmedium' H. S., das mit Rhineschen Kartenversuchen auf mutmaßliche ASW-Leistungen getestet wurde und damals das kleine Eichhalde-Team in Atem hielt, zumal vier IGPP-Mitarbeiter selbst persönlich Zeugen spektakulärer physikalischer Phänomene geworden waren, landläufig als ‚spontane Psychokinese', vulgo ‚Spuk', bezeichnet (siehe Abb. 1).

Abbildung 1

Kartenversuch mit dem Spukmedium H. S. am IGPP unter Beteiligung von Dieter Vaitl (2. von rechts). Quelle: Archiv des IGPP

Abbildungen 2a – 2c

Brief von Dieter Vaitl an Hans Bender zu dessen 80. Geburtstag

JUSTUS - LIEBIG - UNIVERSITÄT GIESSEN

FACHBEREICH 06 PSYCHOLOGIE
Klinische Psychologie - Prof. Dr. D. Vaitl

6300 Gießen, den 15.1.1987
Otto Behaghel-Straße 10
Telefon (0641) 702-5385
Telex 482956

Sehr verehrter, lieber Herr Bender,

zu Ihrem 80. Geburtstag würde ich Ihnen am liebsten das sinnreiche Kapitel 28 aus Musils "Mann ohne Eigenschaften" abschreiben: "Ein Kapitel, das jeder überschlagen kann, der von der Beschäftigung mit Gedanken keine besondere Meinung hat". Als uomo colto, von den Chariten beschenkt, haben Sie mir als Lehrer Türen zu Gedankenwelten aufgestoßen. In einer Ihrer Vorlesungen sprang der Funke Ihrer Begeisterung für Robert Musil auf mich über. Die Beschäftigung mit diesem säkularen Werk hat mich bis heute nicht aus seinem Bann entlassen. Ich danke Ihnen dafür.

Im Juni 1986, als die Deutsche Gesellschaft für Psychophysiologie und ihre Anwendung ihre 15. Tagung in Freiburg abhielt, haben wir uns nach vielen Jahren wiedergesehen. Wie groß war meine Freude, daß Sie eigens zum Vortrag Ihres Schülers gekommen waren. Ich gestehe gern: Ich war sehr aufgeregt, wie es sich gehört, wenn der bewunderte Lehrer sich für das Fachgebiet interessiert, in das einer seiner vielen Schüler abgewandert ist.

In den Gesprächen, die ich an jenen Sommertagen in der "Eichhalde" mit Ihnen und der wunderbaren Lotte Böhringer - bei saftigem Apfelkuchen und trockenem Gutedel - geführt habe, begeisterte mich ein Begriff,

den Sie prägten: die Erosion der Evidenz; von ihr werden Ihrer Meinung nach alle Erlebnisse und unmittelbaren Erfahrungen paranormaler Ereignisse befallen und im Nu zerfressen. Aber: es gibt auch eine "evidentia in se compacta tota", und diese erlaube ich mir für meine Erlebnisse und Erfahrungen zu beanspruchen, die ich als studentische Hilfskraft in Ihrem Institut von 1964 - 1967 gesammelt habe. Während dieser Lehrzeit haben Sie mich an aufregenden Erkundungsfahrten in Bezirke der Parapsychologie teilnehmen lassen. Sie bleiben mir unvergeßlich.

Mit Ihrem merkurischen Gefährt, dem unverwechselbaren DKW, fuhren wir einmal in das Markgräfler Land, um "Einspielungen" auf ihre Echtheit hin zu überpfüfen. Die nächtliche Heimfahrt im Gewitter bildete die rechte Kulisse für die Auseinandersetzungen über die Denkwürdigkeiten, die wir im Hause von Herrn R. gemeinsam erlebt hatten. Waren es tatsächlich Stimmen aus einer anderen Welt, die sich in Radiosendungen eingemischt hatten oder bloße Projektionen eines phantasievollen, polyglotten Menschen, der jedwedes Sprachfragment minutiös mit Bedeutung belegte? Kategorisch habe ich die Projektionslösung verteidigt, während Sie als erfahrener Lehrer diesen casus weit vorsichtiger beurteilten und mich ermahnten, naheliegenden und engen Lösungen stets zu mißtrauen. Die Evidenz dieser Empfehlung ist noch nicht erodiert, ich versuche sie bis heute zu beherzigen.

Ihr Enthusiasmus und Ihre vitale Ausdauer zwangen uns Hilfskräfte damals, durch Rhine'sche Kartenversuche und Psychokinese-Experimente mit einem Medium (H. S.) bis tief in die Nacht hinein dem Zufall etwas abzutrotzen. Sie, lieber Herr Bender, waren nie ein "Apologet des Zufälligen" (Odo Marquard). Immer wieder haben Sie beklagt, daß der Kampf um die Bedeutung von "Statistik", "Antizufallswahrscheinlichkeit" und "Chaos" nicht mit gleicher Schärfe auch in den anderen Disziplinen der Psychologie geführt werde wie gerade auf dem umstrittenen Gebiet der Parapsychologie. Warum soll es nicht in Psychologie **und** Parapsychologie deterministisches Chaos geben? Eine Evidenz hierfür stelle ich mir sehr reizvoll vor, mag sie auch bald wieder - wie bei vielen Meta-Modellen - von Erosion befallen sein.

Ein psychopathologisch interessantes Erfahrungsfeld waren für mich jene klinischen Fälle, die Sie als mediumistische Psychosen bezeichneten und die damals Ihrem Institut zur besonderen Begutachtung überwiesen worden waren. Auch hier halfen Sie mir, die Besonderheit eben dieser Problematik zu erkennen. Mich hat beeindruckt, daß man solange gefeit ist gegen all das, was sich an verwundenen Denkstilen, Weltsichten und Krankhaftem bei diesen Psychosen abspielte, als man die Prämissen dieser Patienten nicht übernimmt. Für den angehenden Psychodiagnostiker war hier die spontane Erosion der Evidenz ein heilsamer Schutz; nur der stete Rückbezug auf die Nicht-Evidenz eben dieser Prämissen entwand mich dem Sog, der vom Argumentationsduktus dieser Patienten ausging. Sie, lieber Herr Bender, halfen mir, beides sowohl bewundernd als auch kritisch im Auge zu behalten: das Kunstwerk ("Gesamtkunstwerk"?) der Gedanken gleich wie das Absurde des ersten Gliedes in diesen perfekt geknüpften Gedankenketten.

Das Gedanken-Kapitel von Musil, glaube ich, zeigt einen Weg, der Erosion von Evidenzen zu entgehen: es muß viel gedacht und es muß Außergewöhnliches gedacht werden. Die Evidenz des Unmittelbaren und Alltäglichen ist eben darum so beglückend, weil sie sich wie ein Parasit an alles heftet, was die All-Gemeinheit liebt und schätzt: "Man muß also sagen, daß ein Mensch, wenn er nur ein bißchen nachdenkt, gewissermaßen in eine recht unordentliche Gesellschaft gerät" (ebendort im 28. Kapitel). Sie haben es gewagt, lieber Herr Bender !

Da Ihnen die Glücksfeigheit vieler älterer Menschen fern ist, wünsche ich nur noch, daß Ihnen in den kommenden Jahren viel Glück zufällt und Sie der Mut zum Ungewöhnlichen nicht verläßt.

Dieter Vaitl

(Dieter Vaitl)

In seiner Ansprache „Neue Entwicklungen in der Spukforschung“ von 1969 als Präsident der „Parapsychological Association“ hat Bender die damaligen Vorgänge geschildert:

> Im März 1966 wurden elektrische Leitungen im Keller eines Schulneubaues [in Freiburg] verlegt. Eine große Anzahl von Haken mussten in den Zementmauern befestigt werden. Mit Schlagbohrer wurden je zwei 8 mm breite Löcher in die Wände getrieben und der Haken dann mittels zweier Schrauben und zweier Plastikdübel befestigt. Der Vorarbeiter beobachtete, dass fast unmittelbar nach der Fixierung der Haken die Schrauben locker wurden. Die Haken, an denen er vorher Klimmzüge machen konnte, ließen sich nun unschwer aus der Wand entfernen. Am folgenden Tag sind nach dem Zeugnis mehrerer Beobachter sieben Haken aus der Zementwand mit den Dübeln herausgekommen. Einer soll dem Lehrjungen, als er durch den Gang ging, in einer Kurve nachgeflogen sein. Wir hatten daraufhin in ‚erwartender Beobachtung‘ ein Experiment veranstaltet, an dem das Team des Freiburger Instituts und einige andere Beobachter teilnahmen. Zwei Haken wurden in der Zementwand befestigt und auf ihre Festigkeit geprüft. Der Lehrjunge stand einen Meter von der Wand entfernt. Wir beobachteten in gespannter Aufmerksamkeit die Schrauben. Innerhalb von zwei Minuten waren sie lose; keiner von uns hat sie herauskommen sehen. Wir dokumentierten mit Blitzlichtaufnahmen und Tonband. Zwei Tage nach diesem erfolgreichen Experiment versuchten wir eine Filmdokumentation, aber dieses Mal blieben die Haken fest.

Kein Wunder, dass diese Beobachtungen als das „Phänomen der lockeren Schrauben von Freiburg“ in die juristisch-kriminologische Literatur eingegangen sind …

II

Das 1996 eingerichtete internationale „Mind/Machine Interaction Consortium: PortREG Replication Experiments“, das auf dem Hintergrund der neuen Forschungsmittel aus der Holler-Stiftung an das IGPP möglich wurde, kam unter wesentlicher Mitwirkung Vaitls und seiner persönlichen Kontakte zu Bob Jahn von der PEAR-Gruppe an der Princeton University zustande. Es bestand aus drei Arbeitsgruppen: (1) der sechsköpfigen PEAR-Gruppe unter Leitung von Robert Jahn, (2) der „Freiburg Anomalous Mind/Matter Interaction“ unter Leitung von Johannes Mischo, Emil Boller und Holger Bösch sowie (3) dem „Giessen Anomalous Research Project“ unter Leitung von Dieter Vaitl selbst, Joop Houtkooper und Bertram Walter. Das Gesamtresultat lieferte in Bezug auf die Wiederholbarkeit der fraglichen Effekte ein heterogenes Bild, stachelte aber Vaitls persönliche Wissbegier (oder Ehrgeiz?) soweit an, dass er seine Wagner-Begeisterung – die er mit Bob Jahn teilte – während der Bayreuther Festspiele über drei Jahre auf die Psi-Probe stellte. Da es – meines Wissens – das einzige experimentelle Projekt auf dem Gebiete der parapsychologischen Forschung geblieben ist, bei dem Vaitl sowohl als Experimentator wie auch als Versuchsperson teilgenommen hat, sei zumindest das Abstract

vollständig wiedergegeben – eine formale Publikation liegt nicht vor. Vorgestellt wurde das Projekt beim „4th Biennial European SSE Meeting in Valencia, Spain", und zitiert ist es nach dem Newsletter der SSE *The Explorer*, Fall 1998, Volume 1, Number 4 (keine Seitenangaben):

Anomalous Effects During Richard Wagner's Operas

Dieter Vaitl
Dept. of Clinical and Physiological Psychology
University of Giessen, Giessen, Germany

Music evokes strong and subtle emotions in man. This is especially true for Richard Wagner's operas which are composed in a sophisticated way (tonal associative network or the "magic of relations") that fits uniquely the duration and the pattern of human responding to emotional material, as has been shown in previous psycho-physiological recordings during his operas in the festival theater at Bayreuth. The annual Wagner festival can also be viewed as a group ritual at a "sacred site" since it constitutes a pilgrimage by devotees to the "Festspielhaus", a theater designed by the composer for optimal resonance with his work. The exploratory Field REG applications reported here were designed to learn more about the circumstances that favour anomalous deviations of a random event generator (REG), as described by the PEAR group in numerous scientific publications. In 1996, 1997, and 1998, data were taken with the Field REG during the first cycle of the following Wagner operas performed at the Bayreuth festival theater: Fliegender Hollaender, Tristan und Isolde, Meistersinger von Nuernberg, Rheingold, Walkuere, Siegfried, Goetterdaemmerung, and Parsifal. The features of the first cycle that foster a quasi-experimental repeated measurement design are the identity of time and place, of conductors and singers, and last but not least, of the majority of listeners, who are Wagner experts rather than novices. The separate acts of the operas were designated as the temporal segments for the experiment. The results reveal a marginally significant effect for the operas in 1996 (Chi-square 13.70, 7 *df*, *P*-value 0.057) and 1997 (Chi-square 27.88, 19 *df*, *P*-value 0.086). In 1998 the outcome was Chi-square 19.91, 16 *df*, *P*-value 0.224. Because the trends are consistent, a formal analysis combining data for 1996, 1997, and 1998, shows a significant effect (Chi-square 61.5, 42 *df*) with *P*-value 0.026. The Bayreuth data sets are compared with Field REG applications by members of the PEAR group during eight other operas and shows in New York City: Brigadoon, Seven Deadly Sins, Carmina Burana, Haensel and Gretel, I Puritani, Faust, Hoffmanns Erzaehlungen, and Russalka, revealing a non-significant overall effect (Chi-square 16.063, 16 *df*) with *P*-value 0.653.

An act-wise inspection of the Bayreuth datasets with the largest contributions to the Field REG deviations reveal a common feature indicating that resonance phenomena (e.g. during choir parts) and highly emotion-arousing scenes might have played a role. Although it seems to be premature, at present, to delineate an explanatory model from these data or to infer meaning from the direction of deviations, the results suggest, in a

pragmatic sense, that the repeated application of the Field REG methodology under nearly identical circumstances, such as in operas, will allow the collection of data which may reveal, in the near future, components involved in man/machine interaction.

III

Die Zusammenarbeit zwischen Dieter Vaitl und mir in den beiden Jahrzehnten im Rahmen des Vorstandes und im Rahmen der Alltagsgeschäfte des Instituts verlief weitgehend geräuschlos. Mitunter hat's – besonders am Anfang – etwas geruckelt, als mich Dieter Vaitl darauf hingewiesen hat, dass das Institut nicht Bauer-Institut heiße (womit er natürlich recht hatte), allerdings auch nicht Vaitl-Institut (womit ich recht hatte), worauf wir uns – nach Reflexionen über ein „Hans-Bender-Institut" HBI – auf die Abkürzung IGPP einigten.

In dem von mir oben erwähnten Brief an seinen akademischen Lehrer Hans Bender registriert Dieter Vaitl, dass diesem – Bender – „die Glücksfeindlichkeit vieler älterer Menschen fern" sei.

Abbildung 3

Dieter Vaitl und der Autor anlässlich der Feier zur Gründung des BION in Gießen 2000 mit einer Fotografie von Hans Bender. Quelle: Archiv des IGPP

Was man sicher auch mit Fug und Recht auch von ihm – Dieter Vaitl – sagen konnte, der auf Komplimente über sein schmuckes Äußeres gerne erwiderte: „Der Erpel zeigt sein Gefieder".

In diesem Sinne ist auch Dieter Vaitls persönliches Abschiedsgeschenk und sein Vermächtnis an das Bendersche Erbe zu verstehen – zur Feier des 70jährigen Gründungsjubiläums des IGPP im Frühjahr 2020 erschien ein von ihm herausgegebener und mitverfasster umfangreicher Forschungsband *An den Grenzen unseres Wissens – Von der Faszination des Paranormalen*.

Meine Lesart von der dort abgedruckten „Frühen Geschichte des IGPP" fand Vaitl so überzeugend, dass er das Kapitel mitunterzeichnete.

Ich darf abschließend den Stoßseufzer zitieren, den ein ehemaliger Präsident der „Parapsychological Association" anlässlich seiner Präsidentenansprache geäußert hat und der da lautet:

> *God in his wisdom created psi*
> *But forgot to tell us why!*

– und diesen entsprechend modifizieren

> *God in his wisdom created Vaitl*
> *But this decision was really vital!*

Literatur

Bauer, E., & Vaitl, D. (2000). Wie alles begann: Frühe Geschichte des IGPP (von den Anfängen bis zu Hans Benders Tod). In D. Vaitl (Hrsg.), *An den Grenzen des Wissens: Von der Faszination des Paranormalen* (S. 27–47). Herder.

Bender, H. (1970). Neue Entwicklungen in der Spukforschung. *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie*, *12*(1), 1–18.

Jahn, R., Dunne, B., Bradish, G., Dobyns, Y., Lettieri, A., Nelson, R., Mischo, J., Boller, E., Bösch, H., Vaitl, D., Houtkooper, J., & Walter, B. (2000). Mind/Machine Interaction Consortium: PortREG Replication Experiments. *Journal of Scientific Exploration, 14*(4), 499–555.

Timm, U., & Vaitl, D. (1967). Zeitschriftenspiegel. *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie*, *10*, 146–155.

Vaitl, D. (1965). J.H. Phillips: Psychoanalyse und Symbolik (Buchbesprechung). *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie*, *8*(1), 126–128.

Vaitl, D. (1965). Helen Pleasants (Editor): Biographical Dictionary of Parapsychology (Buchbesprechung). *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie*, *8*(1), 128.

Zeitschrift für Anomalistik
Journal of Anomalistics

www.anomalistik.de www.igpp.de

„Gleitflug ins Licht"

Nachruf auf Professor Dieter Vaitl (1940–2026)

Ina Schmied-Knittel[a]

Prof. em. Dr. Dieter Vaitl ist am 10. Januar 2026 im Alter von 85 Jahren verstorben. Als ehemaliger Direktor und Vorsitzender des Trägervereins des Instituts für Grenzgebiete der Psychologie und Psychohygiene e. V. (IGPP) war er über zwanzig Jahre mein Chef – einer, der ein vertrauensvolles Arbeitsumfeld schaffte und als Mentor und Förderer agierte. Im Laufe dieser Zeit gab es etliche Anlässe und Projekte, die uns – den Psychophysiologen und die Soziologin – auch als Forschungsteam in den sogenannten Grenzgebieten zusammenbrachten, und ich möchte die Gelegenheit nutzen, Erinnerungen wachzuhalten und Dankbarkeit auszudrücken.

(a) Institut für Grenzgebiete der Psychologie und Psychohygiene
schmied@igpp.de

Professor Vaitl war Pionier auf dem Gebiet der psychophysiologischen Forschung, insbesondere im Bereich Neuroimaging (bildgebende Verfahren). Seine fachliche Expertise reichte weit darüber hinaus, umfasst das große Gebiet der Klinischen Psychologie, Verhaltensmedizin, Emotionsforschung sowie die Erforschung veränderter Bewusstseinszustände wie Meditation, Hypnose, Trance und Halluzination. (Insgesamt listet sein Profil auf *Research Gate* 253 Publikationen auf.)

Es war der letztgenannte Bereich der außergewöhnlichen Bewusstseinszustände, der unser beider Forschungsinteressen zusammenführte, namentlich das Thema Nahtoderfahrung. Ich selbst habe dem Thema meine Anstellung am IGPP zu verdanken – das Ins-

titut finanzierte seinerzeit ein soziologisches Forschungsprojekt zu Nahtoderfahrungen, nach dessen erfolgreicher Beendigung ich als feste Mitarbeiterin übernommen wurde. Das Thema begleitet meine Forschungsarbeit bis heute. Prof. Vaitl kam längst vorher im Rahmen seiner Forschungen zu veränderten Bewusstseinszuständen mit dem Thema in Berührung, primär auf der Grundlage psychophysiologischer und neurodynamischer Fragestellungen und Methoden. Gleichwohl waren ihm die „schwächelnden naturwissenschaftlichen Erklärungsbemühungen" (Vaitl, 2022, S. 49) und reduktionistischen Modelle, denen zufolge alle außergewöhnlichen Erfahrungen ausschließlich mit veränderter Hirnphysiologie zu erklären seien, durchaus ein Dorn im Auge. Vaitl plädierte deshalb auch und gerade im Kontext der Nahtodforschung für mehrdimensionale Untersuchungsansätze, wie er sie mit seinem BION-Team an der Universität Gießen in der Meditationsforschung längst realisiert hatte. Zugleich verwies er auf erkenntnistheoretische Schwierigkeiten und auf diesbezügliche „Grenzen unseres Wissens" (so Vaitls literarisches Vermächtnis über die Wissenschaftsgeschichte des IGPP; Vaitl 2020), insbesondere, wie unbefriedigend die Antworten der Neurologie, Biologie und Psychologie ausfallen: „Man muss feststellen, dass bis heute noch kein neurobiologisches Modell existiert, das die Vielfalt und Verschiedenartigkeit der NTE zufriedenstellend erklären würde", so Vaitls (2022, S. 49) abschließendes Urteil gegenüber seiner eigenen Disziplin. An dieser Stelle erfragte er immer wieder meine soziologische Expertise und zitierte empirische Befunde meiner Nahtod-Publikationen, etwa, dass soziologische Untersuchungen „stets die große Bandbreite von NTE" betont haben und dass ihre Ausgestaltung und Deutung „größtenteils von kulturellen Einflüssen" abhängen (ebd., S. 52).

Die Forderung nach Klärung und Interdisziplinarität war Gegenstand einer Ringvorlesung im *Studium Generale* der Universität Freiburg, die ich gemeinsam mit Dieter Vaitl im Wintersemester 2013 organisierte. Der Titel: „Nahtod-Erfahrungen: Mythos und Realität". (Die anschließenden Abendessen mit dem Opernfreund und Genussmenschen waren legendär.)

In vieler Hinsicht waren die Nahtoderfahrungen für Professor Vaitl ein Paradebeispiel. Ihren realen Stellenwert als persönlich beglaubigte Erfahrungen und die Orientierung an der subjektiven Perspektive, die die Erlebnisse für die jeweiligen Betroffenen haben, war auch für den Hirnforscher zentral. Und selbstverständlich wusste er um das großes Konsistenzbedürfnis bei der Einordnung und Interpretation solcher Erfahrungen und darum, welch hohen Stellenwert sie in der individuellen Lebensgeschichte haben. Auch sein eigenes wissenschaftliches Interesse an Nahtoderfahrungen kann mit Sicherheit nicht losgelöst von seinem persönlichen christlichen Glauben altkatholischer Prägung betrachtet werden, wenngleich er meilenweit davon entfernt war, die Nahtoderfahrung mit einem Jenseits in Zusammenhang zu bringen, wie dies so mancher Nahtodforscher anstrebt. Vaitl dagegen forderte Zurückhaltung und Demut,

überließ als Privatmensch der christlichen Theologie den Vortritt und verwies als Naturwissenschaftler auf die „Grenzen der Erkenntnis“ – so ein weiterer seiner Herausgebertitel im Kontext der Grenzgebietsforschung (Mayer et al., 2015).

Als Herausgeber dieses Handbuchs für Anomalistik zeigt sich auch Professor Vaitls positive und zugewandte Haltung gegenüber den Grenzgebieten und den Mitarbeiter/innen seines Freiburger Instituts. In ganz grundsätzlicher Weise forderte und förderte er entsprechende Publikations- und Forschungsprojekte, auch wenn sie weit weg vom sogenannten Mainstream waren. „High risk“ nannte er diese Unterfangen und plädierte in seiner Rolle als IGPP-Leiter stets auch für die Realisierung von Forschungsprojekten, die methodisch oder inhaltlich unter Umständen ein hohes Risiko an Kontroversen und Rückschlägen bergen.

Bis zuletzt und auch mit schwerer Krankheit war Professor Vaitl dem IGPP und uns Mitarbeitenden verbunden. Ich mochte es, wenn er anrief und mir persönlich für Neuerscheinungen gratulierte oder sich nach geplanten Forschungsprojekten erkundigte: stets höflich, immer interessiert und im Übrigen ein vorzüglicher Smalltalker. Zu seinem traurigen Abschied möchte ich ihm gerne einen „*Gleitflug ins Licht*“ wünschen – treffende Worte, die er selbst einst als Überschrift wählte (Vaitl, 2017). Sein Denken und Wirken werden weit über sein Leben hinaus nachhallen.

Literatur

Mayer, G., Schetsche, M., Schmied-Knittel, I., & Vaitl, D. (Hrsg.). (2015). *An den Grenzen der Erkenntnis. Handbuch der wissenschaftlichen Anomalistik*. Schattauer.

Vaitl, D. (2017). Gleitflug ins Licht. *Herder Korrespondenz, Heft 2,* S. 32–35.

Vaitl, D. (Hrsg.). 2020. *An den Grenzen unseres Wissens*. Herder.

Vaitl, D., Gripentrog-Schedel, S. S., Kugele, J., & Popkes, E. E. (2022). Nahtoderfahrungen. Ein wissenschaftliches Mysterienspiel heute: Interview mit Dieter Vaitl vom 11. November 2017. In *Grenzarbeiten auf der Nulllinie: Nahtoderfahrungen in interdisziplinärer Perspektive*. Universitätsverlag Kiel, S. 47–54. https://doi.org/10.38072/978-3-928794-66-4/p3

Zeitschrift für Anomalistik
Journal of Anomalistics

www.anomalistik.de www.igpp.de

In Memoriam: Stephen Edward Braude (1945–2026) – der Jazzphilosoph *sui generis*[1]

Etzel Cardeña[a]

(a) Lund University, Sweden

Am 3. Januar verlor die Parapsychologie nicht nur ihren bedeutendsten Philosophen der letzten Jahrzehnte, sondern auch einen einzigartigen, vielseitig begabten Geist. Stephen E. Braudes umfangreiches Werk (sieben Bücher und mehr als 100 weitere Veröffentlichungen) zeichnete sich durch prägnante Analysen und deren Implikationen (manchmal mit beißendem Humor gewürzt) sowie durch scharfe

1 Dies ist die deutsche Übersetzung des Nachrufs von Etzel Cardeña, der im *Journal of Anomalous Experience and Cognition*, *6*(1), pp. 7–10, erschienen ist. https://doi.org/10.31156/jaex.28980

Die Übersetzung wurde von Gerhard Mayer mit Unterstützung von DeepL und Google Übersetzer bewerkstelligt.

Kritik sowohl an skeptischen Positionen als auch an einigen gängigen Positionen der Parapsychologie aus. Esteban (wie ich ihn nannte) und ich lernten uns um die Jahrtausendwende kennen, als wir beide in Maryland lebten. Wir teilten nicht nur ein Interesse an Psi, sondern auch an Dissoziation und, wie sich herausstellte, an obskurer Avantgarde-Musik. Ein kurzer Nachruf kann ihm nicht gerecht werden, aber ich werde einige seiner wichtigsten Ideen vorstellen und Quellen auflisten, damit interessierte Leser sich eingehender damit befassen können (siehe auch https://psi-encyclopedia.spr.ac.uk/articles/stephen-e-braude).

Makro-Psychokinese

Braude war jahrzehntelang einer der entschiedensten Verfechter der Makro-Psychokinese – eine Einsicht, die durch die Erfahrung einer Levitation eines Tisches in einer Gruppe genährt wurde. Er führte unter anderem an, dass: 1) die Beweise für Makro-PK überzeugend seien und sich auch nach Abzug der betrügerischen Fälle nicht einfach wegdiskutieren lassen, und dass 2) die Unterscheidung zwischen Mikro- und Makro-PK irreführend sei (Braude, 1986, 2015). Er führte zudem Beobachtungsstudien an physikalischen Medien und außergewöhnlichen Personen durch, wie etwa Katie, der „Gold Leaf Lady", auf deren Körper sich plötzlich dünne Messingfolie bildete und sie teilweise bedeckte, und er widerlegte skeptische Darstellungen der „thoughtographs" [Gedankenfotografien] von Ted Serios (Braude, 2007; 2015). Ich schätze seinen Mut, Phänomene in ihrem jeweiligen Milieu zu untersuchen – etwas, das Forscher zu Beginn der Parapsychologie meisterhaft praktizierten, heute aber nur noch selten anzutreffen ist.

Mögliches Überleben des Todes

Bei der Erörterung von Belegen für ein Weiterleben nach dem Tod, wie sie sich aus verschiedenen Evidenzbereichen wie etwa des mentalen Mediumismus ergeben, argumentierte Braude (2003), dass stichhaltige Belege aus diesen Quellen nicht einfach wegdiskutiert werden können, und wog verschiedene Psi-Erklärungsansätze gegeneinander ab. In erster Linie verglich er die „Living-Agent-Psi"-Hypothese (oder „Super-Psi"-Hypothese), der zufolge die Informationsquelle von lebenden Personen stammt, mit der Hypothese, dass die Quelle in verstorbenen Personen zu suchen ist. Seine sorgfältigen Argumente werden in einem seiner Bücher auf über 300 Seiten dargelegt und weisen einfache, pauschale Antworten zurück: „Wir können – wenn auch mit geringer Gewissheit, aber doch mit einer gewissen Berechtigung – sagen, dass die vorliegenden Belege eine annehmbare Grundlage für den Glauben an ein persönliches Weiterleben nach dem Tod bieten. Sie stützen nicht eindeutig die Überzeugung,

dass jeder den Tod überlebt: Vielmehr stützen sie eher die Überzeugung, dass dies einigen gelingt … für eine begrenzte Zeit" (Braude, 2003, S. 306). Seine Schlussfolgerung ähnelt jener des führenden Forschers auf dem Gebiet des mentalen Mediumismus, Alan Gauld (1983). Bei beiden Autoren schätze ich, dass sie die Beweise und Argumente darlegen, die ihre (moderaten) Schlussfolgerungen stützen, im Gegensatz zu jenen lautstarken Autoren, die es als axiomatisch ansehen, dass ein Weiterleben ohne funktionierendes Gehirn unmöglich sei, oder dass die vorhandenen Beweise schlüssig belegen würden, dass unsere Persönlichkeiten/Identitäten über den Tod hinaus fortbestehen.

Weitere Themen

Steve diskutierte auch Dissoziation, insbesondere das, was früher als *multiple Persönlichkeitsstörung* bezeichnet wurde (heute: *dissoziative Identitätsstörung*; Cardeña, 2017), sowie Hypnose, dissoziative Prozesse und persönliche Einheit; dabei setzte er sich kritisch mit den Leugnern von Dissoziation und der Möglichkeit wiedererlangter Erinnerungen auseinander (Braude, 1991). Hinsichtlich parapsychologischer Hypothesen stand er den experimentellen Belegen mitunter ablehnend gegenüber und vertrat die Ansicht, dass Psi im realen Leben weitaus stärker ausgeprägt sei (wobei er der PK Priorität einräumte; Beispiele für anomale Kognition finden sich jedoch etwa bei Stevenson, 1970). Eine Lektion von ihm, die in den letzten Jahren in der Community nicht in ausreichendem Maße gelernt wurde, besagt, dass mentale Phänomene anderer Natur sind und nicht auf dieselbe Weise untersucht oder verstanden werden sollten wie physikalische.

Steve äußerte sich mit Nachdruck über die Unehrlichkeit vieler ignoranter und dogmatischer Anti-Psi-Kritik (Beispiele hierfür finden sich bei Cardeña, 2015) und fand zudem kurz angebundene Worte für Psi-Erklärungen, die seiner Meinung nach wenig zur Klärung beitrugen. Zu Letzteren zählen retrokausale Erklärungen der Präkognition (1986) und mechanistische Erklärungen für parapsychologische (und psychologische) Phänomene (Braude, 2014).

Trotz der Intoleranz gegenüber Psi wurde Steve emeritierter Professor und ehemaliger Lehrstuhlinhaber für Philosophie an der University of Maryland Baltimore County sowie Chefredakteur des *Journal of Scientific Exploration*. Er erhielt zahlreiche Stipendien, Förderungen und Auszeichnungen, darunter den Distinguished Achievement Award der International Society for the Study of Trauma and Dissociation und die F. W. H. Myers Memorial Medal der Society for Psychical Research.

Steve Braude: Der ganze Mensch

So beeindruckend seine philosophischen Beiträge zur Parapsychologie auch sind, es steckte (oder steckt?) noch weitaus mehr in ihm (Braude, 2020). Stets elegant gekleidet, schuf Steve nicht nur Bücher und Fachartikel, sondern auch Musik. Er bezeichnete sich selbst als „Jazz-Philosophen", da er Jazz-Piano spielte (Aufnahmen einiger seiner Stücke mit der Yamara Band sind auf Spotify und Apple Music zu finden oder können unter https://www.jazzphilosopher.com/ gestreamt werden; Noten seiner Kompositionen lassen sich zudem unter https://www.jazzphilosopher.com/scores herunterladen). Er war auch sehr bewandert in klassischer Musik, und wenn ich einmal eine seltene, vergriffene Aufnahme nicht auftreiben konnte, durfte ich oft darauf zählen, dass er sie für mich aufnahm und mir zusandte.

Steve war furchtlos und handelte eher aus Prinzip als aus Eigeninteresse. Vor einigen Jahrzehnten, noch bevor wir tatsächlich gute Freunde geworden waren, wurde ich von Adrian Parker und seinem Verbündeten Göran Brusewitz angegriffen, weil ich die Stelle bekam, auf die sich Parker ebenfalls beworben hatte. Im Laufe von mehr als 10 Jahren beinhalteten ihre Feindseligkeiten den Versuch, meine Aufnahme in eine geschlossene Parapsychologie-Mailingliste zu verhindern, die falsche Darstellung des Themas der Doktorarbeit einer meiner Doktorandinnen gegenüber einem Prüfer, woraufhin dieser aus dem Prüfungsausschuss austrat, sowie eine Beschwerde bei der Parapsychological Association wegen einer Auszeichnung, die diese mir verliehen hatten. Inmitten eines ohrenbetäubenden Schweigens von vielen in der parapsychologischen Community, die sich lediglich im privaten Rahmen auf meine Seite stellten, unterstützte mich Steven öffentlich, was ich ihm ungemein hoch anrechnete.

Zum Abschluss möchte ich Steves Komplizin Djurdjina „Gina" Ruk (ebenfalls stets elegant) mein tiefstes Beileid aussprechen.

Literatur

Braude, S. (1980). *ESP and psychokinesis: A philosophical examination*. Temple University Press.

Braude, S. (1986). *The limits of influence: Psychokinesis and the philosophy of science*. Routledge & K. Paul.

Braude, S. (1991). *First person plural: Multiple personality and the philosophy of min*d. Rowman & Littlefield.

Braude, S. (2003). *Immortal remains: The evidence for life after death*. Rowman & Littlefield.

Braude, S. (2007). *The gold leaf lady and other parapsychological investigations*. University of Chicago Press.

Braude, S. (2014). *Crimes of reason: On mind, nature and the paranormal*. Rowman & Littlefield.

Braude, S. (2015). Macro-psychokinesis. In E. Cardeña, J. Palmer, & D. Marcusson-Clavertz (Hrsg.), *Parapsychology: A handbook for the 21st century* (pp. 258–267). McFarland

Braude, S. (2020). *Dangerous pursuits: Mediumship, mind, and music*. Anomalist books.

Cardeña, E. (2015). The unbearable fear of psi: On scientific censorship in the 21st century. *Journal of Scientific Exploration, 29*(4), 601–620.

Cardeña, E. (2017). Dissociation and the dissociative disorders. In J. Stein (Hrsg.), *Reference module in neuroscience and biobehavioral psychology*. Elsevier.

Gauld, A. (1983). *Mediumship and survival*. Flamingo.

Stevenson, I. (1970). *Telepathic impressions: A review and report of thirty five new cases*. University of Virginia Press.

Zeitschrift für Anomalistik
Journal of Anomalistics

www.anomalistik.de www.igpp.de

Nachruf auf P. Andreas Resch (1934–2026)[1]

Peter Mulacz[a]

Andreas Resch © Cincelli

(a) Österreichische Gesellschaft für Parapsychologie und Grenzbereiche der Wissenschaften (ÖGPP)
office@parapsychologie.ac.at

Pater DDr. Andreas Resch C.Ss.R (geb. 29.10.1934 in Steinegg, Südtirol) war von 1969 bis 2000 Professor für *Klinische Psychologie und Paranormologie* an der Accademia Alfonsiana der Päpstlichen Lateranuniversität/Päpstliche Universität Gregoriana in

1 Der Nachruf erschien zuerst im Newsletter N° 90 der Österreichischen Gesellschaft für Parapsychologie und Grenzbereiche der Wissenschaften vom 29. März 2026.

Rom, 1978 Gründer des von ihm bis 1980[2] geleiteten *Instituts für Grenzgebiete der Wissenschaft* (IGW) in Innsbruck und 1974 des Resch-Verlags, wo er der Herausgeber mehrerer Zeitschriften und Buchreihen war; hervorzuheben ist dabei die Zeitschrift *Grenzgebiete der Wissenschaft* (GW), welche die Fortführung der von dem katholischen Verleger Josef Kral 1951 gemeinsam mit Abt Wiesinger[3] initiierten Zeitschrift *Verborgene Welt* darstellt (wobei dann u. a. P. Mag. Ferdinand Zahlner, Nachfolger von P. Dr. Peter Hohenwarter in der Leitung der [längst erloschenen] *Arbeitsgemeinschaft für Parapsychologie* an der *Wiener Katholischen Akademie*, als Mitherausgeber fungierte). Alle Nummern der Zeitschrift GW von 1951 bis zum Ende ihres Erscheinens können nunmehr kostenlos heruntergeladen werden.[4]

Weiters war er Initiator der IMAGO MUNDI-Kongresse, die zunächst an verschiedenen Orten, dann – nach dem Ausbau des IGW-Gebäudes – immer in Innsbruck abgehalten wurden, zeitweise alternierend mit den *Basler Psi-Tagen*. Von den Imago Mundi-Kongressen sind 15 Jahresbände erschienen. Unter den Referenten dieser Kongresse finden sich viele bekannte Namen, z. B. W. H. C. Tenhaeff, Gerda Walther, Gabriel Marcel, Hubert Larcher, Gebhard Frei, Erwin Nickel, Ernst Benz, Georg Sigmund, Werner Bonin, Wihelm Gauger u. v. a. m.

Es wäre zu kurz gegriffen, nur die organisatorischen Aspekte von Reschs Wirken in seinem Verlag und im Kongressbetrieb zu erwähnen; selbst verfasste er mehrere Bücher und zahlreiche Aufsätze, viele davon thematisch an der Schnittstelle von Parapsychologie und Religion bzw. Theologie angesiedelt, wozu ihn sein Doppelstudium der Psychologie und der kath. Theologie prädestiniert hatte. Die Themen beinhalteten u. a. Wunder von Seligen und Heiligen, Kanonisationsverfahren, die Erscheinungen am Wallfahrtsort Medjugorje und andere Marienerscheinungen, Geister, Sindone und Schleier etc.

Die von ihm verwendete Begrifflichkeit weicht vom Üblichen ab. Er bezeichnete die Disziplin in ihrer Gesamtheit nicht als Parapsychologie, sondern als Para*normologie*, wobei dieser Oberbegriff die Gebiete Paraphysik, Parabiologie, Parapsychologie und Parapneumatologie umfasst; seine Vorstellung vom Wesen der menschlichen Seele hat die Aspekte Psyche (*ψυχή*) und Pneuma (*πνεῦμα*) unterschieden – das war ihm ein besonderes Anliegen. Seine Terminologie hat sich aber nicht durchgesetzt.

2 1980 erfolgte die Umwandlung in einen gemeinnützigen Verein, als solches besteht das IGW nach wie vor. Der URL für das IGW incl. GW und aller anderen Aktivitäten lautet: https://www.imagomundi.biz/

3 Wiesinger O. Cist., A. (1948): *Okkulte Phänomene im Lichte der Theologie.* Styria.

4 https://www.imagomundi.biz/gw/

Ein Anliegen war es ihm auch, an die *Psychotronik* – Bezeichnung der Parapsychologie in den damaligen Warschauer Pakt-Staaten – anzudocken, wie dies als erstem wohl Stanley Krippner gelungen ist. Für die 1973 gegründete *Internationale Gesellschaft für Psychotronische Forschung* (IAPR) gab er die Zeitschrift *Psychotronik* heraus, die allerdings nach der dritten Nummer ihr Erscheinen einstellen musste.[5]

Es soll nicht unerwähnt bleiben, dass Resch ob seiner Befassung mit Parapsychologie bzw. Grenzgebieten auch kirchlicher Widerstand entgegenschlug, was er unbekümmert ertragen hat.

Als Priester wirkte er jahrzehntelang im Redemptoristenkloster von Innsbruck; seinen Lebensabend verbrachte er im Redemptoristenkloster von Gars am Inn in Bayern, wo er am 01. Februar 2026 verstarb und wo er am dortigen Klosterfriedhof seine letzte Ruhe gefunden hat.

5 Download möglich: https://www.imagomundi.biz/psychotronik/

Zeitschrift für Anomalistik
Journal of Anomalistics

www.anomalistik.de www.igpp.de

Nachruf auf Josef Dvořák (1934–2026)[1]

Peter Mulacz[a]

Josef Dvorak © Helmut Wolech

Josef Dvořák (geb. 28.01.1934 in Wien) war eine interessante und vielschichtige Persönlichkeit, eine gewisse Exzentrizität ist ihm nicht abzusprechen. Seine vielfältigen Aktivitäten machen es schwer, ihn mit *einem* Wort – z. B. einer Berufsbezeichnung – zu kennzeichnen; vielleicht hilft eine Auflistung: Schriftsteller, Journalist (u. a. beim legendären *Forvm*), Psychoanalytiker, Performance-Künstler, aber begonnen hat er als Theologe (Schüler von Karl Rahner).

(a) Österreichische Gesellschaft für Parapsychologie und Grenzbereiche der Wissenschaften (ÖGPP)
office@parapsychologie.ac.at

1 Der Nachruf erschien zuerst im Newsletter N° 90 der Österreichischen Gesellschaft für Parapsychologie und Grenzbereiche der Wissenschaften vom 29. März 2026.

Als Religionswissenschaftler galt sein Interesse den Sekten und Okkultgruppen, insbesondere dem Satanismus, über den er ein Buch[2] veröffentlich hat, das heute als Standardwerk gilt; nachdem er auch ein entsprechendes Ritual öffentlich vorgeführt hat, wurde er vielfach als „Satanspriester" diffamiert: eine Unterstellung, die geradezu grotesk ist.

Dvořák war wohl der Erste, der Carl Kellner der Vergessenheit entrissen hat und mit ihm den Okkultismus des Fin de Siècle in Wien, dessen Protagonisten Friedrich Eckstein und der genannte Carl Kellner waren, jeweils mit den Schwerpunkten Theosophie bzw. Yoga. Dieser spezifische Wiener Okkultismus um 1900 ist mittlerweile gut erforscht: Hier ist vor allem Karl Baier, vormals Professor am Institut für Religionswissenschaft der Universität Wien, zu nennen, der darüber ausführlich publiziert hat:

- Baier, K. (2018). Yoga within Viennese Occultism: Carl Kellner and Co. In K. Baier, P.A. Maas, & K. Preisendanz (Hrsg.), *Yoga in Transformation. Historical and Contemporary Perspectives* (S. 387–438). V & R unipress.
- Baier, K. (2020). Occult Vienna: From the Beginnings until the First World War. In A. Mattes, H. G. Hödl, & L. Pokorny (Hrsg.), *Religion in Austria*, *5*(5), 1–76.

Prof. Baier hat auch in unserem Rahmen einen thematisch einschlägigen Vortrag gehalten: Der Wiener Okkultismus des Fin de Siècle, 30. Oktober 2023.

Zum „Psychoanalytiker" Dvořák (die Anführungszeichen beziehen sich auf die nicht schulgemäße, „wilde" Art seiner Praxis): Seine Lehranalyse hat Dvořák bei Wilfried Daim absolviert.

Exkurs zu Wilfried Daim (1923–2016):

> Daim war Psychologe; in den Jahren nach dem Ende des Zweiten Weltkriegs, noch während seines Studiums, führte Daim privat Experimente zur Telepathie in Träumen durch, wobei er farbige Kartonstücke unterschiedlicher Formen als Objekte benutzte, die dann von der Versuchsperson in ihre Träume (verändert) eingebaut wurden – z. B. verwendete Daim als Objekt eine kleine runde orange Pappscheibe, und die Träumerin träumte von einer orangen Sonnenscheibe in einem Bild von Odilon Redon, oder es wurde ein grünes Dreieck als Zielobjekt im Traum zu einem Tannenbaum verfremdet. (Publikation: *Experimente mit der Seele. Naturwissenschaft für jedermann,* Graz 1949.)
>
> Später war Daim eine Zeit lang Vizepräsident der Österreichischen Gesellschaft für Psychische Forschung (wie unsere Gesellschaft vor der Namensänderung 1971 hieß).

2 Dvorak, J. (1989). *Satanismus. Schwarze Rituale, Teufelswahn und Exorzismus, Geschichte und Gegenwart*. Heyne.

> Er nahm auch (als einziger Österreicher) an dem von der Parapsychology Foundation (PF) organisierten I. Internationalen Kongreß für Parapsychologie in Utrecht (1953) teil. Sein Lehranalytiker war Igor Graf Caruso gewesen, allerdings ging Daim sehr bald eigene Wege (Publikation: *Umwertung der Psychoanalyse*, Wien 1951).
>
> Politisch war Daim ein sogenannter „Linkskatholik"; es war überraschend zu sehen, dass es Daim trotz seiner psychoanalytischen Ausbildung nicht möglich war, sich von einer sehr reaktionären repressiven katholischen Sexualmoral zu emanzipieren, wie sich anlässlich einer Diskussion über Diefenbachs monumentalen Fries *Per aspera ad astra* gezeigt hat.
>
> Später trat Daims Aktivität als Kunstsammler und -Händler in den Vordergrund, insbesondere als Experte für O. R. Schatz. Daim war auch imstande, Hans Scheibner als den Schöpfer der holzgeschnitzten Säulen am Entrée des Café Landtmann in Wien zu identifizieren.

Dvořák hat als Psychoanalytiker jene Künstler behandelt, die dann als Exponenten des Wiener Aktionismus berühmt geworden sind. Dessen Ursprung lag in einem von Hermann Nitsch, Adolf Frohner und Otto Muehl unterzeichneten Manifest *Die Blutorgel*, eine Zeitschrift, von der nur vier Nummern erschienen sind und deren Herausgeber ein gewisser „Fritz Graf" war (eines der Pseudonyme, die Dvořák benutzte).

Dvořák betrieb eine Zeitlang auch eine eigene Kunst-Galerie sowie als Veranstaltungsraum den Perinetkeller[3]. Legendär ist die tagelange „Einmauerung" der Künstler im Perinetkeller, um ohne äußere Ablenkung sich ihren künstlerischen Impulsen hingeben zu können. Natürlich war das ein bloßer PR-Gag; nach erfolgter publikumswirksamer Vermauerung der Tür von außen haben die Künstler den Keller durch den Hinterausgang verlassen.

Mittlerweile ist der Wiener Aktionismus als Kunstrichtung etabliert, und seit einigen Jahren gibt es das (mittlerweile deutlich vergrößerte) *Wiener Aktionismus Museum* (WAM-Wien), nunmehr unter der Leitung von Klaus-Albrecht Schröder (bis 2024 Direktor der Albertina).

Halten wir fest: Unter all den vielfältigen Aktivitäten Josef Dvořáks liegt seine Relevanz für unseren Interessensbereich darin, das „okkulte Wien" von vor mehr als einem Jahrhundert „wiederentdeckt" und in den Fokus des Interesses der Kulturwissenschaft gebracht zu haben.

Leider waren seine letzten Lebensjahre traurig, er verbrachte sie – ans Bett gefesselt – in einem Pflegeheim, wobei sein Zustand starken Schwankungen unterworfen war. Gerne hätte ich ihn noch einmal vor seinem Tod gesehen, aber das war mir leider nicht vergönnt.

Josef Dvořák ist am 24. Jänner 2026 verstorben.

3 Der Perinetkeller existiert nach wie vor, derzeit vom *Institut ohne direkte Eigenschaften* (IODE) unter dem Slogan „Aufführungsort der Wiener Aktionisten in den 1960ern, die hier die Kunst des Schönen, Wahren und Guten durch radikale Irritation ersetzten" betrieben, und zwar unter der Leitung von Erich Félix Mauthner.

Zeitschrift für Anomalistik
Journal of Anomalistics

www.anomalistik.de www.igpp.de

Buchrezension

Dean Radin

The Science of Magic

How the Mind Weaves the Fabric of Reality[1]

New York: Harmony, 2025
ISBN 13: 978-0-593-79758-7, 304 Seiten

Rezensent:

Gerhard Mayer [a]

(a) Institut für Grenzgebiete der Psychologie und Psychohygiene
mayer@igpp.de

Der berühmt-berüchtigte und einflussreiche Okkultist Aleister Crowley (1875–1947) definierte in seinem Buch *Magick in Theory and Practice* (Crowley, 1970) Magie folgendermaßen: „Magick is the Science and Art of causing Change to occur in conformity with Will" [Magie ist die Wissenschaft und Kunst, Veränderungen in Übereinstimmung mit dem Willen herbeizuführen] (S. xii).

Das neue Buch von Dean Radin trägt den Titel *The Science of Magic*. Dieser scheinbar einfache Titel hat verschiedene Bedeutungsebenen und -facetten. Er könnte sich auf die wissenschaftliche Untersuchung von Magie als einer Praxisform (aus einer etischen [Außen-]Perspektive) oder auf Magie als einen spezifischen quasi-

1 Dies ist eine deutsche Übersetzung der englischen Originalversion, die auf der Webseite der *Society for Psychical Research* erschienen ist: https://www.spr.ac.uk/ScienceofMagic

wissenschaftlichen Ansatz zum Verständnis und zur „Herbeiführung von Veränderungen" in der materiellen Realität (aus einer emischen [Innen-]Perspektive – in Übereinstimmung mit Crowleys Definition) beziehen, oder „Magie" könnte als alternativer Begriff für „Parapsychologie" stehen. Die letztgenannte Bedeutung mag für einige Parapsychologen herausfordernd sein; sie wurde jedoch bereits von Dean Radin in seinem letzten Buch *Real Magic* (Radin, 2018) eingeführt und entspricht genau seinem Anliegen. Dort schrieb er: „Ich beschäftige mich seit etwa vierzig Jahren aus wissenschaftlicher Perspektive mit Magie. In den ersten neununddreißig Jahren hätte ich diese Aussage vehement bestritten" (S. 11, Übersetzung G. M.). Das neue Buch kann als Variation oder Fortsetzung von *Real Magic* (und seinen anderen früheren Büchern) angesehen werden; es fügt jedoch einen neuen Aspekt hinzu, der über den Bereich wissenschaftlichen Experimentierens hinausgeht.

The Science of Magic ist ein typisches Buch von Dean Radin. Wie alle seine vorangehenden Werke könnte es auch den Untertitel tragen: Eine Einführung in die experimentelle Parapsychologie. Und wie alle seiner früheren Bücher ist es elegant geschrieben und leicht zu lesen; und es richtet sich nicht in erster Linie an Kollegen, sondern an eine breite Leserschaft. Das Buch profitiert von Radins jahrzehntelanger Erfahrung als einem Experimentator, der nicht seine gesamte Forschungskarriere der Verfeinerung einer einzigen Forschungsrichtung oder der Verfolgung einer einzigen geliebten Theorie gewidmet hat, sondern flexibel Vorschläge anderer Forscher einbezogen und gleichzeitig aus dem Scheitern einiger seiner eigenen Experimente gelernt hat.

Doch warum ein neues Buch veröffentlichen, wenn ähnliche Inhalte bereits publiziert sind? Wie schon angedeutet, kann das Buch als Fortsetzung oder Weiterentwicklung früherer Werke und der darin vermittelten Ideen angesehen werden. Die bereits in *The Conscious Universe* (Radin, 1997) vorgestellte Grundthese, dass ein materialistisches und reduktionistisches Weltmodell unzureichend ist und dass Bewusstsein nicht auf das Gehirn reduziert werden kann, bildete die Grundlage seiner Forschung und führte ihn zunehmend weg von der Vorstellung, dass die Lösung auf der Ebene der Physik zu finden sei. Seine eigenen Erfahrungen mit Bewusstseins-Techniken wie Meditation veranlassten ihn unter anderem, paranormale Phänomene in anderen Kulturen, insbesondere in Indien, zu untersuchen und aus der Perspektive der westlichen Wissenschaft zu bewerten (Radin, 2013). Mit seinem Buch Real Magic wandte er sich dann der westlichen Kulturgeschichte zu und stellte Verbindungen zwischen magischen Praktiken und Psi-Phänomenen her, wie sie aus der parapsychologischen Forschung bekannt sind. Diesen Ansatz verfolgt und intensiviert er in seinem neuen Buch, indem er Magie nicht mehr aus der Perspektive der westlichen Wissenschaft betrachtet, sondern Wissenschaft aus der Perspektive der westlichen Magie. Durch die Verwendung entsprechender Kapitelüberschriften nimmt er

ein Reframing vor. Diese Vorgehensweise mag einigen seiner Kollegen irritierend erscheinen, ist aber in sich logisch konsistent.

The Science of Magic ist in vier Teile gegliedert, die sich in ihrer Länge stark unterscheiden. Teil I trägt den Titel „A Revolution is Brewing". Er besteht nur aus dem Einführungskapitel, in dem beschrieben wird, wozu das Buch dient, wozu es nicht dient, wie er „Magie" versteht („I use the terms *magic* and *psi* interchangeably", S. 11)[2] und welche Rolle das Bewusstsein dabei spielt. Der umfangreichste Teil ist der zweite, der wie das Buch selbst den Titel „Science of Magic" trägt und drei Kapitel umfasst: „Science of Enchantment", „Science of Divination" und „Science of Theurgy".

„Science of Enchantment" befasst sich mit verschiedenen experimentellen Ansätzen zum Nachweis der Auswirkungen gezielter mentaler Intention auf der materiellen Ebene, d. h. mit dem, was auch in der Tradition der klassischen Magie und Zauberei berichtet wird. Dies bezieht sich auf verschiedene Formen von PK-Experimenten, die von der Beeinflussung des Verhaltens von RNGs (Zufallszahlengeneratoren), des Zellwachstum, von Photonen, Plasma, Heilmitteln bis hin zur Veränderung der Wasserqualität reichen.

Das Kapitel „Science of Divination" behandelt Experimente zu Phänomenen, die sonst mit den klassischen Begriffen Hellsehen, Präkognition, Retrokognition und Fernwahrnehmung bezeichnet werden. Schließlich befasst sich die „Science of Theurgy" mit wissenschaftlichen Bestrebungen, die sich mit Fragen zum Überleben nach dem physischen Tod und, allgemeiner, zu einer göttlichen oder jenseitigen Sphäre beschäftigen (survival research). Obwohl die Frage nach dem Göttlichen oder Übernatürlichen nicht direkt durch die Wissenschaft beantwortet werden kann, ist es dennoch möglich, zu indirekten Plausibilitätsurteilen zu gelangen – vergleichbar mit der Kosmologie, wo die Konzepte der dunklen Materie oder der dunklen Energie nur indirekt aus den Auswirkungen auf das, was für uns sichtbar oder messbar ist, erforscht werden können.

Teil III, „Understanding Magic", enthält zwei Kapitel, von denen das erste den Titel „Taking Magic Seriously" trägt und sich mit skeptischen Reaktionen und der Tabuisierung der Psi-Forschung befasst („... despite perpetual fascination by billions of people, psi as real is willfully ignored by over 99 percent of the academic and scientific world" – S. 163). Zusätzlich zu der großen Mehrheit der Forschenden, die die Beweise ignorieren – von denen wohl die meisten einfach nicht das notwendige Interesse an den großen Fragen mitbringen, die die Parapsycho-

2 Eine solche Gleichsetzung erscheint mir etwas unglücklich, da ich Magie als eine Form des Handelns verstehe, die auf bestimmten Überzeugungen basiert, während Psi ein Begriff ist, der verwendet wird, um eine bestimmte Klasse von *Phänomenen* zu beschreiben.

logie für ein besseres Verständnis der Welt aufwirft –, kämpft eine kleine Gruppe ideologischer Skeptiker aktiv gegen die Anerkennung der wissenschaftlichen Erkenntnisse der Parapsychologie. Dabei folgen diese einem dreiteiligen Narrativ, wie Radin schreibt:

> Dismiss anecdotal evidence as unreliable, loudly assert that experimental evidence that seems to support psi or magic is solely due to flaws or fraud, and then write off these phenomena as impossible because they supposedly violate physical laws [Anekdotische Beweise als unzuverlässig abtun, lautstark behaupten, dass experimentelle Beweise, die Psi oder Magie zu stützen scheinen, ausschließlich auf (methodischen) Mängeln oder Betrug beruhen, und diese Phänomene dann als unmöglich abtun, weil sie angeblich gegen physikalische Gesetze verstoßen] (S. 169).

Nach einer langen Phase der Entzauberung der Welt durch die Wissenschaft ist laut Radin in den letzten Jahren jedoch eine „Wiederverzauberung" zu beobachten, die neben dem Ernstnehmen persönlicher subjektiver Erfahrungen auf einem differenzierteren Verständnis der Bedingungen menschlicher Existenz basiert. Der entscheidende Faktor auf wissenschaftlicher Seite ist in erster Linie die Bewusstseinsforschung (consciousness research), die in den letzten Jahren erheblich an Dynamik gewonnen und auch einen starken Einfluss auf die Theoriebildung hat.

Kapitel 6 befasst sich mit „Theorien der Magie" – zunächst aus einer emischen Perspektive, d. h. wie magisch Praktizierende ihre Praxis theoretisch verstehen und welche Bedingungen erfüllt sein müssen, damit diese erfolgreich ist. Radin skizziert dann kurz die grundlegenden Unterschiede zwischen zentralen philosophischen Weltanschauungen (Materialismus, Dualismus, Panpsychismus, Duale-Aspekte-Monismus), bevor er schließlich fünf Klassen wissenschaftlicher Theorien zur Magie vorstellt.

Der Titel von Teil IV, „Practical Magic", weist auf jenen Aspekt hin, der dieses Buch deutlich von seinen Vorgängern unterscheidet, nämlich dass es Anleitungen für das praktische Ausüben von Magie anbietet. Dean Radin stellt mehrere klassische Methoden wie Sigillenmagie, Knotenmagie und Kerzenmagie vor. Er gibt auch Vorschläge, wie man das „Funktionieren" von Magie selbst leicht testen kann. Auf den letzten Seiten des Buches (Kapitel 12, „Beyond Magic") beantwortet er Fragen, die häufig in Frage-Runden zu diesem Thema aufkommen, wie etwa, warum die Effekte in parapsychologischen Experimenten, aber auch in der magischen Praxis üblicherweise sehr klein sind, worin der Nutzen der Magie bestehen kann usw.

Man kann davon ausgehen, dass vor allem dieser Teil IV des Buches – trotz der geschickten Vorbereitung in den ersten drei Teilen – Kritik sowohl von Kollegen als auch von Skeptikern hervorrufen wird. Letztere wissen seit jeher, dass Forschende, die sich der Parapsychologie widmen, an magisch-irrationalem Denken leiden und als seriöse Wissenschaftler:innen nicht

ernst genommen werden können; und erstere möchten vielleicht den Schritt von der Theorie zur Praxis nicht gehen, weil ihnen die Gleichsetzung von Magie und Psi zu gewagt erscheint und sie vor allem nicht des magisch-irrationalen Denkens bezichtigt werden wollen. Ich muss gestehen, dass auch ich zunächst etwas zurückhaltend gegenüber der Idee war, dass das Buch einen praktischen Leitfaden enthalten würde. Nach der Lektüre des Buches waren meine Vorbehalte jedoch zerstreut, da Radin die Leserschaft in überzeugender Weise zur Praxis hinführt und deren instrumentellen Charakter als Teil von Bewusstseins-Techniken betont.

Die Kritik an Radins früheren Büchern (Radin, 1997, 2006, 2013), nämlich dass er skeptischen Stimmen zu wenig Raum gibt und bestehende Probleme in der parapsychologischen Experimentalforschung nicht angemessen behandelt, lässt sich auch auf dieses Buch übertragen. In Anbetracht seines Anliegens sowie der Zielgruppe und der unerbittlichen Angriffe ideologischer Skeptiker wäre es jedoch möglicherweise zu viel verlangt, auf alle internen kritischen Debatten hinzuweisen. Mit seiner jahrzehntelangen Erfahrung in der experimentellen Parapsychologie und seiner ausgezeichneten Kenntnis der Fachliteratur sind seine Argumente fundiert.

Als Rezensent sollte ich versuchen, eine Antwort auf die berechtigte Frage zu geben, ob das Buch für Parapsycholog:innen lesenswert ist, die bereits mit der Forschung vertraut sind und sich normalerweise nicht für populärwissenschaftliche Werke interessieren. Ich sehe zwei Gründe, es dennoch zu lesen:

1. Der Überblick über die vielen interessanten Experimente mit positiven Ergebnissen, betrachtet aus einer vereinheitlichenden Perspektive der Wechselwirkung zwischen Geist und Materie und der Bewusstseinsforschung, bietet ein erfrischendes Gegenmittel gegen die oft übermäßig kritische und selbstzerstörerische interne Kritik unter Parapsychologen, die bei ihren Bemühungen um die Dekonstruktion ihres Forschungsgebiets offenbar erfolgreicher sind als manche Skeptiker.

2. Radins Perspektivwechsel und Neudefinition können Kolleginnen und Kollegen dazu ermutigen, vertraute Standpunkte auf ähnliche Weise zu hinterfragen. Man kann beispielsweise neue Ideen zu den Vorgehensweisen besonders erfolgreicher Experimentatoren, zur elusiven Natur der Resultate bei Psi-Experimenten und zu den oben erwähnten kleinen Effektstärken von Psi entwickeln, ohne auf axiomatische One-size-fits-all-Lösungen zurückgreifen zu müssen, die von quantenphysikalischen Modellen vorgeschlagen werden und einen einfachen Ausweg aus den komplexen Bedingungen für das Auftreten von Psi-Phänomene versprechen.

Referenzen

Crowley, A. (1970). *Magick in theory and practice / by the Master Therion*. Castle Books.

Mencken, H.L. (1921). *Vorurteile: Zweite Reihe*. Jonathan Cape. http://archive.org/details/prejudices030184mbp

Radin, D. (1997). *Das bewusste Universum: Die wissenschaftliche Wahrheit über psychische Phänomene*. HarperEdge.

Radin, D. (2006). *Entangled minds: Extrasensory experiences in a quantum reality*. Paraview.

Radin, D. (2013). S*upernormal: Wissenschaft, Yoga und der Beweis für außergewöhnliche psychische Fähigkeiten*. Deepak Chopra Books.

Radin, D. (2018). *Echte Magie: Alte Weisheit, moderne Wissenschaft und ein Leitfaden zur geheimen Kraft des Universums*. Harmony Books.

Zeitschrift für Anomalistik
Journal of Anomalistics

www.anomalistik.de www.igpp.de

Buchrezension

Corey Lee Wrenn

Vegan Witchcraft: Contemporary Magical Practic and Multispecies Social Change

London: Routledge, 2025
ISBN: 978-1032649719, 222 Seiten, € 50,00

Rezensentin:

Meret Fehlmann[a]

(a) Institut für Sozialanthropologie und Empirische Kulturwissenschaft, Universität Zürich

Dass *Human-Animal-Studies* (HAS) und *Pagan Studies* zusammenfinden, war nur eine Frage der Zeit, in Anbetracht der vor bald 20 Jahren eingesetzten Blüte der beiden Forschungsfelder. Beide Richtungen sind von einer ähnlichen Mischung aus Aktivismus und Spiritualität geprägt und Teil des vielgestaltigen Diskurses rund um das Anthropozän. Mit Corey Lee Wrenns *Vegan Witchcraft* liegt nun eine erste Monografie vor, die sich aus soziologischer Perspektive spezifisch diesem Nexus widmet. Die Studie gliedert sich in sieben Kapitel, um diesen Verbindungen auf den Grund zu gehen.

Bevor die Autorin in *Introduction to Vegan Witchcraft* die Grundlagen für ihre Studie legt, informiert sie über ihre eigene Verbundenheit als Hexe und Veganerin zu ihrem Forschungsfeld (S. X). In der Einleitung arbeitet sie heraus, dass in ihrem Verständnis Hexerei und Veganismus über weite Strecken identische Einstellungen teilen (oder teilen sollten), da sie „interrogate humanity's domination over nature" (S. 1). Veganismus wie moderne Hexerei entstammen teil-

weise den gleichen Gegen-/Subkulturen und arbeiten sich an den dichotomen Vorstellungen zum Verhältnis von Natur und Kultur, Mann und Frau, Mensch und Tier ab. Wie sie schreibt, stünde es beiden Richtungen gut an, sich mehr aufeinander und die darin verbreiteten theoretischen Paradigmen einzulassen (vgl. S. 2–7).

Veganismus wird in der englischsprachigen Forschung teilweise als atheistisch aufgefasst, während in der deutschsprachigen Forschung eher von einer prägenden spirituellen oder auch religiösen Komponente ausgegangen wird (vgl. Funkschmidt, 2016). So oder so herrscht im Veganismus die Vorstellung einer belebten Natur vor, außerdem hat diese Bewegung Wurzeln im Anarchismus des 19. Jahrhunderts mit seinem Fokus auf Unterdrückungsmechanismen. Gemäß Wrenn seien bisher vor allem die drei monotheistischen Religionen und asiatische Religionen auf ihren Umgang mit nichtmenschlichen Tieren hin untersucht worden, ein „critical animal approach to pagan religions" (S. 7) stehe noch aus, wofür sie mit ihrer Arbeit steht.

Das zweite Kapitel *What is Witchcraft* befasst sich mit der Geschichte der Hexerei. Auch wenn sie darauf hinweist, dass die historischen und heutigen Hexen nicht gleichzusetzen seien, dass aktuelle Hexenbilder vor allem Ausdruck der Projektionen gegenwärtiger Befindlichkeiten seien, rapportiert sie eine Geschichte der Hexerei in einer tendenziell oberflächlichen Art und Weise. Neue Perspektiven ergeben sich in der Auseinandersetzung mit dem (Spannungs-)Verhältnis zwischen Wicca und Veganismus (vgl. S. 16).

Das nächste Kapitel fokussiert auf *Animal Familiars*, also auf die insbesondere im britischen Hexenglauben verbreitete Vorstellung eines Hilfsgeistes der Hexe meist in Tiergestalt (vgl. zu diesem Aspekt z. B. Kapitel 10 in Hutton, 2017). Dieses Zusammenleben von Hexe und *familiar* sowie die Fähigkeit der Hexe, sich in Tiere zu verwandeln, versteht die Verfasserin als Ausdruck der Unterwanderung der Vorstellung der menschlichen Suprematie, indem die Grenzen zwischen nichtmenschlichen und menschlichen Tieren (wie in HAS die Differenz von Menschen und Tieren sprachlich aufgehoben wird, vgl. dazu Kompatscher-Gufler et al., 2017) durchlässig werden. Wrenn schreibt: „It seems a difficult task for pagans to engage with other animals without reducing them to tropes and spirits to be absorbed or attempting to physically control their body and mind" (S. 43).

Das anschließende Kapitel befasst sich mit *Green Witchcraft* als Ausdruck ökofeministischen, paganen Aktivismus' zur Versöhnung von Mensch, Mitgeschöpfen und Natur. Als charakteristisch für *green witchcraft* sieht Wrenn die „collaboration with nature, with its rejection of domination and owner ship, distinguishes green witchcraft from patriarchal religions" (S. 59). In diesem Mindset findet eine Gleichsetzung von Hexe, Natur und Erde statt. Ökofeminismus hat eine große Nähe zum spirituellen Feminismus und gewissen Spielarten neopaganer

Religion, es geht um die Überwindung der Spaltung des Spirituellen vom Materiellen. Kultur ist in diesem Verständnis nach Wrenn ein Produkt und als solches formbar, woraus sich ihr zufolge ein Auftrag ableitet: „There is an element of creativity here, imagining the unknown, inter acting with other animals and nature, and forging a culture that is symbiotic and egalitarian" (S. 64).

In Kapitel 4 wendet sich Wrenn der *Kitchen Witchery* zu, gerade auch traditionell bestand eine enge Verbindung von Hexen zu Lebensmitteln, denn die Zubereitung, aber auch Haltbarmachung von Essen galt als einfachste Art von Magie (vgl. die Studie von Purkiss, 1996). Wrenn geht davon aus, dass der Veganismus, indem er mehr als eine Ernährungsform ist, physische und spirituelle Veränderungen auszulösen vermöge (vgl. S. 96). Die im Neuheidentum generell verbreitete Tendenz, die Vergangenheit als golden zu betrachten, findet sich besonders stark in der *kitchen witchery* wieder. In neuheidnischen Kreisen ist die Vorstellung verbreitet, mittels angepasster Ernährungsweise eine Nähe zu den imaginierten Vorfahren:innen aus alter (heidnischer) Zeit zu erreichen, wobei auf „amateur research into anthropology, folklore, archaeology, or earlier topical publications in witchcraft, and providing recipes and practices that reflect these imaginations to provide the practitioner a point of connection" (S. 102) zurückgegriffen wird.

Der aktivistische Charakter wird durch Wrenns Ausführungen oder vielmehr Werturteile immer wieder deutlich, wenn sie auf den Vorschlag einer Hexe in einem Kochbuch, möglichst lokal und saisonal zu kochen und zu essen, um sich so das Land einzuverleiben, mit einem solchen Fazit reagiert: „What might non-vegan witches become upon ingesting places of slaughter and suffering?" (S. 101).

Wrenn kritisiert vor allem das weit verbreitete Lob für Milchprodukte, sie sieht diese als Ausdruck der Ausbeutung der nichtmenschlichen Tiere, zudem deutet sie diese Ernährungsform als unreif, da das Abstillen ein natürlicher Schritt in der kindlichen Entwicklung sei. Mindestens so gravierend ist in ihrer vegano-aktivistischen Haltung der Umstand, dass der Konsum von Milchprodukten ein westliches, koloniales Ernährungsprinzip sei, das andere Ethnien, die kein Milcheiweiß verdauen können, pathologisiere. Tierprodukte sind nach Wrenn (symbolisch) mit den männlichen und wohlhabenden Klassen verbunden, und ihr Verzehr, ja bereits der Vorschlag, dass man diese konsumieren könne, sei der Versuch, an dieser ausbeuterischen Praxis teilzunehmen (vgl. insbesondere S. 101–105).

In *Sabbats and Specisism* liegt der Fokus auf der rituellen Praxis. Die rituellen Feiern im Neuheidentum zu den Tag- und Nachtgleichen sind mit Sonnen- und Mondzyklen verbunden, aber letztlich Ausdruck von zutiefst „archaic speciesist practices" (S. 124). Um relevant zu bleiben, weisen die meisten Religionen eine gewisse Flexibilität auf, der synkretistische Charakter neopaganer Religionen erlaubt die Integration von Veganismus, der Ausschluss von Speziesismus sollte keine negativen Folgen für die Authentizität haben.

Wurden praktisch ausnahmslos in den bisherigen Kapiteln die Vorteile des Veganismus besungen, zieht Wrenn in *Spellwork for a Vegan World* Parallelen zwischen Hexerei und (politischem, veganem) Aktivismus, da beide der Befreiung aus Zwängen verpflichtet sind. Trotz ihrer Kritik an „traditioneller" neopaganer Religiosität ist Hexerei ein Mittel zum sozialen Aufbruch: „Witchcraft has thus faced significant political limitations, but it maintains a powerful ethic of resistance and inclusiveness" (S. 152). Unabhängig von der politischen Einbettung ziele Hexerei auf „creating change" (S. 154).

Viele Veränderungen zielen auf Selbstoptimierung und wirken individualistisch. Zugleich gelten sie Wrenn vor dem Hintergrund jahrhundertelanger patriarchaler Unterdrückung von Frau (und Natur) als erster Schritt zur Stärkung weiblichen Selbstwerts und Selbstbewusstseins. Feminismus und Ökologie gehen Hand in Hand, um die jüdisch-christlich-patriarchale Weltordnung zu überwinden (S. 157).

Wrenn arbeitet auch heraus, dass Hexerei wie Veganismus in esoterischen und anarchistischen Milieus des 19. Jahrhunderts Vorläufer:innen haben. In dem Zusammenhang erwähnt sie Zsuzsanna Budapests nur Frauen vorbehaltenen *Susan B. Anthony Coven*. Dessen Benennung erscheint mir jedoch als eine Volte der Geschichte: Plausibler wäre die Bezugnahme in der Namensgebung auf Matilda Joslyn Gage, die sich – anders als die Quäkerin Susan B. Anthony – von zentralen Strömungen des Feminismus entfernte und sich verstärkt esoterischen Interessen zuwandte (zur Geschichte der vergessenen Matilda Joslyn Gage vgl. Brammer, 2000). Hier wie auch in ihren Ausführungen zu WITCH bleibt Wrenn hinter dem etablierten Wissen stehen (vgl. Eller, 1995, Erstausgabe 1993).

In ihrer *Conclusion* betont Wrenn erneut die Nähe von Veganismus und Hexerei, die beide auf eine Verbesserung der Welt abzielen, moniert aber bestehendes Nachbesserungspotenzial der Hexerei in Sachen Speziesismus: „In practice, however, I have argued that witchcraft systematically fails to realize this responsibility to the self, society, and even nature" (S. 183). Der modernen Hexerei ist ein Paradox inhärent, wenn weibliche Spiritualität und Natur als wertvoll angesehen werden, dabei aber Tierwohl nur in ungenügender Weise beachtet wird. *Vegan Witchcraft* befasst sich nach Wrenn mit „three important components in the reproduction of shared oppression between feminized human groups and other animals" (S. 194).

Wie deutlich wurde, bin ich mit dem Buch nicht so richtig warm geworden, das fängt schon mit dem Titelbild an, das den Kopf eines Ouessant-Widders zeigt, dessen Ohr mit Ohrenmarken gekennzeichnet ist, also seine Unterdrückung im ausbeuterischen System der menschlichen Tiere anzeigt. Weshalb dieses Sujet gewählt bzw. von der Autorin durchgewunken wurde, kann ich nicht nachvollziehen, da ihre Hauptargumentation gerade auf den verwerflichen Charakter

der Tierproduktion abzielt und man bei Routledge bekanntlich für ein „individuelles“ Titelblatt extra bezahlt.

Ein grundsätzlicheres Problem, das ich mit dem Buch habe, ist, dass es mir mehr aktivistisch als wissenschaftlich scheint. Es ist zwar bereits Allgemeinwissen, dass unsere Interessen, Sozialisation und Erfahrungen sich in der wissenschaftlichen Arbeit ausdrücken, dass man sich seiner Positionierung im Forschungsfeld bewusst sein soll (vgl. Chakkalakal, 2026), aber im vorliegenden Fall verdrängt der vegane Furor oft die kritische Distanz zum Gegenstand. Wahrscheinlich ist diese Nähe zum Feld bei Forschungen zu Neopaganismus und Veganismus praktisch unvermeidlich, da beides Felder sind, die von aktivistischen Positionen durchdrungen sind, was von den Forschenden eine Extraportion an Reflexivität und methodischer Selbstverortung verlangt, um die eigene Involviertheit transparent zu machen und so analytisch produktiv werden zu lassen. Sonst besteht das Risiko, dass vor allem das Sendungsbewusstsein prägend ist. So ist hier immer wieder dem Aktivismus mehr Raum und Gewicht gegeben als einer vertieften und sich die Mühe einer Kontextualisierung machenden wissenschaftlichen Beschäftigung (als Beispiel sei nur erwähnt, dass sie auf S. 38 von Sibyl Leek spricht, ohne deren Bedeutung für den Hexendiskurs in der Mitte des 20. Jahrhunderts zu erläutern).

Dennoch unternimmt Wrenn mit *Vegan Witchcraft* erste Schritte, um den neuen Hexenglauben aus einer vom Antispeziesismus geprägten Perspektive kritisch zu beleuchten.

Literatur

Brammer, L. R. (2000). *Excluded from suffrage history: Matilda Joslyn Gage, nineteenth-century American feminist*. Greenwood Press.

Chakkalakal, S. (2026). Ethnografische (Auto)Fiktion und der unauflösbare Widerspruch von Immersion und Distanz. In F. Faust, A. Mazukatow, K. Nagel & T. Todd Sekuler (Hrsg.), *Queer-ethnografische Potenziale* (S. 169–182). Barbara Budrich. https://doi.org/10.3224/84743152.11

Eller, C. (1995). *Living in the lap of the Goddess: The feminist spirituality movement in Am*erica. Beacon Press.

Funkschmidt, K. M. (2016). Glauben mit dem Gaumen: Religiöse Ausläufer des Veganismus. *Forum Erwachsenenbildung, 49*(4), 22–25. https://doi.org/10.25656/01:25267

Hutton, R. (2017). *The witch: A history of fear, from ancient times to the presen*t. Yale University Press.

Kompatscher-Gufler, G. P. D., Spannring, R. D., & Schachinger, K. (2017). *Human-Animal Studies: Eine Einführung für Studierende und Lehrende*. utb.

Purkiss, D. (1996). *The witch in history: Early modern and twentieth-century representations*. Routledge.

Zeitschrift für Anomalistik
Journal of Anomalistics
www.anomalistik.de www.igpp.de

Abstracts-Dienst / Literaturspiegel

Der Abstracts-Dienst/Literaturspiegel der *Zeitschrift für Anomalistik* will kurz aktuelle Aufsätze mit Relevanz für die Anomalistik vorstellen, die in herkömmlichen akademischen (d. h. natur-, sozial-, geistes- und kulturwissenschaftlichen sowie medizinischen) Fachzeitschriften erschienen sind. Die Auswahl der zusammengefassten Arbeiten erfolgt stichprobenartig und ist als pragmatischer Literaturspiegel gedacht, will also keinen Anspruch auf Vollständigkeit oder Repräsentativität erheben. Dennoch soll diese kleine Rubrik zwei wichtige Funktionen erfüllen: Zunächst ist sie als Informationsservice für unsere Leser gedacht, die keinen oder nur beschränkten Zugriff auf akademische Zeitschriften haben. Darüber hinaus soll sie die Rezeption von anomalistischen Themen im wissenschaftlichen Mainstream dokumentieren und somit eine kontinuierliche Standortbestimmung der Anomalistik dort vornehmen, wo wissenschaftliche Erkenntnis laufend verhandelt wird: innerhalb der internationalen Fachzeitschriftenlandschaft.[1]

Frauke Schmitz-Gropengießer & Gerhard Mayer

Acunzo, D. J., Denton, A. H., Weiler, M., & Kelly, E. F. (2026). Neuroimaging studies of extrasensory perception: A systematic and critical review. *NeuroImage, 333*, 121937. https://doi.org/10.1016/j.neuroimage.2026.121937

[Neuroimaging-Studien zur außersinnlichen Wahrnehmung: Eine systematische und kritische Übersicht]

Zusammenfassung: In Forschungsarbeiten, die sich über mehr als sieben Jahrzehnte erstrecken, wurde mittels funktioneller Neurobildgebung untersucht, ob mutmaßliche außersinnliche Wahrnehmung (ASW/rezeptives Psi) nachweisbare neuronale Korrelate aufweist. Dem Forschungsfeld fehlt jedoch eine kohärente und kritische Zusammenfassung seiner methodischen Ansätze und der berichteten Effekte. Wir erstellten eine systematische Übersichtsarbeit, in der

1 Wir möchten die Leser:innen einladen, uns auf potenzielle Kandidaten für den Abstracts-Dienst der *ZfA* aufmerksam zu machen. Vorschläge für geeignete Beiträge aus wissenschaftlichen Fachzeitschriften bitte per E-Mail an mayer@anomalistik.de. Unser besonderer Dank gilt Ulrich Magin und Harald Grauer, die uns mit einigen diesbezüglichen Hinweisen versorgt haben.

143 Berichte zusammengefasst und die Methoden von 129 Einzelstudien qualitativ bewertet wurden. Wir haben die Studien anhand der zugrunde liegenden Paradigmen in zwei große Kategorien eingeteilt: (1) explizite Psi-Paradigmen (einschließlich der Unterkategorien „Forced-Choice"- und „Free-Response"-Design), bei denen Psi anhand offener Reaktionen erfasst wird; und (2) implizite Psi-Paradigmen (einschließlich der Unterkategorien „Fernstimulation", „Fernintentionalität" und „prädiktive Antizipationsaktivität"), bei denen Psi ausschließlich anhand neurophysiologischer Messungen bewertet wird. Die meisten Arbeiten stützten sich auf das EEG (91 %), gefolgt von der fMRT (5 %). Wir identifizierten wiederkehrende methodische Einschränkungen, insbesondere kleine Stichprobengrößen, eine unzureichende Kontrolle bei multiplen Vergleichen und mangelnde analytische Flexibilität. Explizite Paradigmen zeigten selten ein über dem Zufallsniveau liegendes Verhalten, dennoch führten die Forscher häufig neuronale Analysen durch, was auf nicht explizit gemachte Änderungen in der operationalen Definition von Psi hindeutet. Bei impliziten Paradigmen wurden zwar häufig Psi-konsistente Effekte berichtet, die Ergebnisse waren jedoch heterogen und ließen sich nur selten replizieren. Insgesamt sind endgültige Schlussfolgerungen über die neuronalen Korrelate von ESP noch verfrüht. Dennoch identifizierten wir potenzielle Indikatoren – wie die Alpha-Band-Power in Forced-Choice-Designs und eine target-bezogene negative Slow Wave in ereignisbezogenen Designs – als überprüfbare Kandidaten für zukünftige Forschung. Wir geben 13 methodische Empfehlungen zur Förderung kumulativer Fortschritte sowie sechs Empfehlungen für zukünftige Forschungsrichtungen.

Händén-Svensson, S. (2025). The lake and its monster: Communicating connections with landscape, belonging and sense of place with Great Lake Monster narratives. *ARV: Nordic Yearbook of Folklore 2025, 81*, 112–119. https://doi.org/10.61897/arv.81.48343

[Der See und sein Monster: Verbindungen zu Landschaft, Zugehörigkeit und Ortsverbundenheit vermitteln durch Erzählungen über Monster in großen Seen]

Zusammenfassung: Hintergrund: Dieser Artikel untersucht die langjährige Verbindung zwischen einer lokalen Variante eines weitverbreiteten kulturellen Phänomens – See- und Wasserschlangen – und der Gemeinschaft, die sich darum kümmert. Es werden Legenden und Beobachtungsberichte, sogenannte „Memorates", über diese Seeschlange ab dem 17. Jahrhundert analysiert. Die Studie erörtert ferner, wie diese Erzählungen – die die bekannteste schwedische Variante von Wasserschlangen darstellen, nämlich das Ungeheuer des Sees Storsjö (Great Lake Monster)[2] in

2 Dank an Ulrich Magin, der mit der folgenden Anmerkung einer Fehlübersetzung zuvorgekommen ist: „The Great Lake Monster ist das Ungeheuer des Sees Storsjö (= großer See), kein großes Seeungeheuer" (Magin, per E-Mail).

der Provinz Jämtland – als Bindeglied zwischen Mensch, Geschichte, Landschaft und gesellschaftlichem Wandel fungieren und wie sie ein Gefühl der Ortsverbundenheit vermitteln. Diese engen Verbindungen machen das Great Lake Monster zu einem Schutzgeist der Region, zu einem Genius loci.

Ausgehend von der zentralen Frage „Auf welche Weise bringen die Erzählungen über das Great Lake Monster Verbindungen zwischen Menschen, Geschichte, Landschaft und Ort zum Ausdruck?" zeigt der Artikel auf, wie frühe Deutungen dieses Phänomens durch ein Leben geprägt waren, das im alltäglichen Umgang mit Natur und Landschaft verwurzelt war. Da die älteren Legenden eine prägende Grundlage für die spätere Vorstellung des Wesens als Kryptid oder „verborgenen Tieres" bildeten, vertrete ich die These, dass sie als Brücke zwischen der auf dem Runenstein von Frösö (Jämtland, Schweden) dargestellten Vorstellung einer Wasserschlange und der kryptozoologischen Tradition eines im See gesichteten Tieres betrachtet werden können.

Herrero, N. L., Corfdir, Y., Vázquez-Chenlo, A. A., Capurro, L., & Forcato, C. (2025). Exploratory study of non-ordinary states of consciousness during sleep show distinct electrophysiological features from wakefulness and canonical sleep stages. *Scientific Reports*, *15*, 33586. https://doi.org/10.1038/s41598-025-18748-7

[Explorative Studie zu außergewöhnlichen Bewusstseinszuständen während des Schlafs zeigt elektrophysiologische Merkmale, die sich deutlich von denen des Wachzustands und der klassischen Schlafstadien unterscheiden]

Zusammenfassung: Das Bewusstsein erlischt während des Schlafs nicht immer. Stattdessen kann es in Form von außergewöhnlichen Bewusstseinszuständen (non-ordinary states of consciousness – NOSC) wieder auftauchen, wie etwa bei luziden Träumen (LD), Schlafparalyse (SP), außerkörperlichen Erfahrungen (AKE) und falschem Erwachen (false awakening – FA). Während einige dieser Zustände bereits phänomenologisch untersucht wurden, sind ihre neurophysiologischen Grundlagen nach wie vor unklar. In dieser explorativen Studie untersuchen wir ihre elektrophysiologischen Korrelate und grenzen sie von den herkömmlichen Schlafstadien ab. Wir führten bei Personen, die diese Phänomene häufig erleben, eine nächtliche Polysomnographie durch und erfassten dabei 10 Episoden (3 LD, 2 SP, 2 AKE, 3 FA). Anhand von Augenbewegungsmarkern wurden Phasen der Luzidität identifiziert. Die relative spektrale Leistung wurde mittels Hauptkomponentenanalyse (PCA) und permutationsbasierter multivariater Varianzanalyse (PERMANOVA) analysiert. Unsere Ergebnisse deuten darauf hin, dass sich diese NOSC vom Wachzustand unterscheiden, jedoch Merkmale sowohl mit Schlafstadium 1 (S1) als auch mit dem REM-Schlaf (Rapid Eye Movement) gemeinsam haben.

Kofman, K., & Levin, M. (2025). Cases of unconventional information flow across the mind-body interface. *Mind and Matter, 23*(1), 13-69. https://doi.org/10.5376/mm2025.13

[Fälle von unkonventionellem Informationsfluss über die Geist-Körper-Schnittstelle]

Zusammenfassung: Die Neurowissenschaften und, im weiteren Sinne, die Verhaltenswissenschaften versuchen, die Beziehung zwischen funktionaler Kognition und den zugrunde liegenden Prozessen in lebendem Gewebe zu beschreiben. Das gegenwärtige Paradigma konzentriert sich stark auf das Gehirn und auf spezifische Mechanismen, von denen angenommen wird, dass sie mentalen Inhalten und Fähigkeiten zugrunde liegen. Einer der interessantesten Ansätze in jedem Forschungsgebiet, der oft zu Fortschritten führt, besteht darin, Daten hervorzuheben, die sich nicht ohne Weiteres in ein bestimmtes vorherrschendes Erklärungsmodell einfügen lassen. Hier geben wir einen Überblick über klinische und experimentelle Daten aus mehreren unkonventionellen Systemen, die von den aktuellen Modellen des Fachgebiets nicht vorhergesagt werden. Eine verringerte Gehirnmasse oder fehlendes Hirngewebe ohne den erwarteten Funktionsverlust (z. B. Hydrozephalus, Hemihydranenzephalie), Diskrepanzen zwischen kognitivem Zustand und Gehirnfunktion (z. B. unbeabsichtigtes Bewusstsein während der Anästhesie, terminale Luzidität) sowie Fälle, in denen die kognitiven Fähigkeiten die offensichtlichen Fähigkeiten des Individuums übersteigen – all dies verdeutlicht interessante Aspekte der enormen Plastizität der Zuordnung zwischen Kognition und ihrem biologischen Substrat. Diese Fälle eröffnen neue Forschungsansätze, die bestehende Rahmenkonzepte zumindest erweitern, und weisen Parallelen zu Entdeckungen auf, die hinsichtlich der entstehenden Form und des Verhaltens synthetischer Konstrukte gemacht werden. Wir entwerfen eine mögliche Forschungsagenda für die Untersuchung interessanter und bislang noch wenig verstandener Merkmale verkörperter Bewusstseinszustände, die für die Biomedizin und ihre Technologien ebenso bedeutsam sein könnten wie für grundlegende philosophische Fragen.

Naish, D. (2026). Monsters of the deep. *Geology Today*, *42*, 87–93. https://doi.org/10.1111/gto.70015

[Monster der Tiefsee]

Zusammenfassung: Seemonster werden meist mit vorwissenschaftlichen Weltbildern oder mit der Fiktion in Verbindung gebracht. Dennoch haben verschiedene Autoren – darunter auch einige Wissenschaftler – Seemonster als real interpretiert und dabei häufig einen Zusammenhang zu Tieren hergestellt, die als Fossilien bekannt sind. Versuche, Seeungeheuer im wissen-

schaftlichen Mainstream zu etablieren, sind weitgehend gescheitert, doch bleiben sie fest in Kultur und Vorstellungswelt verankert. Eine kürzlich in der Aberdeen Art Gallery gezeigte Ausstellung mit dem Titel Monsters of the Deep beleuchtete die Geschichte der Seemonster und zeigte auf, wie wissenschaftliche Entdeckungen und Populärkultur zusammengewirkt haben, das Thema lebendig zu halten.

Scatterty, K. R., VonStein, D., Prichard, L. B., Franczak, B. C., Hamilton, T. J., & Schmaltz, R. M. (2025). Infrasound exposure is linked to aversive responding, negative appraisal, and elevated salivary cortisol in humans. *Frontiers in Behavioral Neuroscience, 20*, 1729876. https://doi.org/10.3389/fnbeh.2026.1729876

[Die Infraschall-Exposition steht beim Menschen im Zusammenhang mit aversiven Reaktionen, einer negativen Bewertung und erhöhten Cortisolwerten im Speichel]

Zusammenfassung:

Einleitung: Als Infraschall werden Schallwellen mit Frequenzen unter 20 Hz bezeichnet, die für den Menschen in der Regel nicht wahrnehmbar sind. Einige Tiere nehmen Infraschall wahr und zeigen eine Abneigung dagegen, was Bedenken hinsichtlich seiner potenziellen schädlichen Auswirkungen als anthropogene Umweltbelastung aufkommen lässt. Jüngste Forschungsergebnisse deuten darauf hin, dass auch Menschen auf Infraschall reagieren könnten, obwohl dieser unterhalb der herkömmlichen Grenze des menschlichen Hörvermögens liegt. Die vorliegende Studie untersuchte die nicht-auditorischen Auswirkungen von Infraschall auf die Stimmung und die Stressreaktion des Menschen.

Methoden: Die Teilnehmenden (n = 36) wurden im Rahmen eines 2×2-Between-Subjects-Designs (beruhigende vs. beunruhigende Musik; Infraschall an vs. aus) entweder beruhigender oder beunruhigender Musik ausgesetzt, wobei Infraschall (~18 Hz) entweder vorhanden war oder fehlte. Selbstauskunftsdaten wurden unmittelbar nach der Exposition erhoben, und Speichelproben wurden unmittelbar vor der Exposition sowie 20 Minuten nach Beginn zur Cortisolbestimmung entnommen.

Ergebnisse: Die Teilnehmenden konnten den Infraschall nicht über das Zufallsniveau hinaus wahrnehmen (p = 0,241). Infraschall war – unabhängig von der Musikbedingung und ohne Erwartungseffekte – mit einem erhöhten Speichel-Cortisolspiegel (p = 0,022, r_{rb} = 0,390) sowie einer stärkeren subjektiven Reizbarkeit (p = 0,049, η^2 = 0,096), Desinteresse (p = 0,044, η^2 = 0,121; p = 0,047, η^2 = 0,118) und Traurigkeitsempfinden (p = 0,002, η^2 = 0,253) verbunden. Mittels Random-Forest-Modellierung wurden Interesse, Reizbarkeit, Traurigkeitsempfinden und Cortisolspiegel zudem als wichtige Prädiktoren für die Infraschall-Exposition identifiziert.

Diskussion: Ohne auditive Wahrnehmung oder Erwartungseffekte war die Infraschall-Exposition mit einem erhöhten Cortisolspiegel und negativeren affektiven Selbstauskünften verbunden. Diese Ergebnisse stimmen mit früheren Tierstudien überein und legen nahe, dass Infraschall auf den Menschen aversiv wirken, als potenzieller Umweltreiz fungieren und zu einem negativeren subjektiven Empfinden beitragen kann.

Tomlinson, M. (2026). Voice and agency in Spiritualist mediumship: or, who speaks when spirits speak? *Journal of the Royal Anthropological Institute*, *32*, 224-243. https://doi.org/10.1111/1467-9655.70008

[Stimme und Agency im spiritistischen Mediumismus – oder: Wer spricht, wenn Geister sprechen?]

Zusammenfassung: Im Spiritismus rücken Medien im selben Akt ihre eigene Agency (Handlungsmacht) sowie jene der Wesenheiten aus der Geisterwelt rituell in den Vordergrund und zugleich in den Hintergrund. Spiritistische Medien übernehmen Verantwortung für die Äußerungen, die sie für ihr Publikum hervorbringen: die Worte, die sie formulieren oder übermitteln, die Heilung, die sie vermitteln, sowie die Schriften und gemalte Bilder, die durch ihre Hände in die Welt gelangen. Gleichzeitig lehnen sie jedoch jede Verantwortung ab, indem sie diese den Wesenheiten in der Geisterwelt zuschreiben; die Zeichen, die Medien über ihren Geist und Körper empfangen, sowie die Botschaften und die heilende Energie, die sie zu vermitteln suchen, sollen aus Quellen stammen, die außerhalb ihrer selbst liegen. Dieses Muster der wechselseitigen Hervorhebung und Zurücknahme von Agency ist historisch beständig, auch wenn sich die Stimme des Mediums insofern gewandelt hat, als sie nicht mehr die Eigenschaften der durch sie sprechenden Entität imitiert, sondern sich durch einen zwanglosen, konversationellen Ton auszeichnet. Die Betrachtung der metaphorischen Verlagerung von Agency im Kontext der Ausdruckshandlungen der Medien gibt Aufschluss darüber, wie diese beim Publikum das Empfinden kultivieren, dass sie zwar nicht die eigentlichen Sprecher sind, aber dennoch niemand sonst auf die gleiche Weise sprechen könnte wie sie.

Guidelines for Authors

- Manuscripts submitted for publication must meet the usual scholarly standards and requirements. Particular attention is paid to logical consistency and clarity of reasoning, and to the use of literature and other documents supporting arguments or hypotheses.
- Manuscripts must be submitted in electronic form in a common text format (*no* pdf files!).
- The manuscript must include an abstract describing as precisely and shortly as possible the content of the article (e.g., questions, methods, results, conclusions). It also must include some keywords.
- Tables and figures must be attached on separate sheets and numbered consecutively. They must be made available in reproducible, good printing quality. For each illustration or table, inscriptions, explanations and, if applicable, source references must be attached on a separate sheet.
- Footnotes are allowed but should be used sparingly.
- Literature references follow the APA (7th edition) style. References must be made in the main text in author-date format. The DOIs of the references must be added if they have one. For precise citation, please consult the stylesheet provided on the Gesellschaft für Anomalistik web pages: https://www.anomalistik.de/en/zeitschrift-fuer-anomalistik/guidelines-for-authors/guidelines-for-authors-of-articles
- The authors grant the *Journal of Anomalistics* / *Zeitschrift für Anomalistik* a simple, non-exclusive, worldwide, and perpetual right to publish it under the Creative Commons License Attribution-Noncommercial 4.0 (CC BY-NC, https://creativecommons.org/licenses/by-nc/4.0/). However, when an article has been submitted to the *JAnom* / *ZfA* for consideration, the Journal holds first serial (periodical) publication rights. Additionally, the publishers have the right to post the article on the Internet and to make it available via electronic as well as print subscription.
- By submitting an article, the author affirms that they hold all necessary copyrights and usage rights to the submitted work and that its publication does not violate any rights of third parties. If the article contains previously published third-party content (e.g., figures, tables, or text excerpts), the author confirms that all necessary permissions for use and publication have been obtained. The authors indemnify the editors and the publisher against any claims by third-parties that arising from copyright or usage rights violations related to the publication of the article.
- All submitted manuscripts are subject to an double-blind peer-review process. Manuscripts that are acceptable in principle but do not quite meet the required standards will either be returned for revisions, and/or edited to be published following the author's approval of edits.
- In order to promote scholarly discussion, in addition to anonymous peer review the editors will usually invite critical comments on a manuscript accepted for publication from other experts. Authors should assume that such commentaries will be published together with their article, and will have the opportunity for author responses.

Hinweise für Autorinnen und Autoren

- Zur Publikation eingereichte Manuskripte müssen den in der Wissenschaft üblichen Standards und Anforderungen entsprechen, insbesondere wird auf logische Klarheit der Argumentation sowie auf die Anführung von Belegen für aufgestellte Thesen geachtet.
- Manuskripte müssen in einem gängigen Textformat (*kein* pdf!) eingereicht werden.
- Dem Manuskript muss ein deutscher und englischer Abstract beigefügt sein. Dem englischen Abstract ist ein englischer Titel für den Gesamtaufsatz zuzuordnen.
- Tabellen und Abbildungen müssen *gesondert* beifügt werden und durchnummeriert sein. Sie müssen in reproduzierbarer, guter Qualität in einem gängigen Format zur Verfügung gestellt werden. Zu jeder Abbildung oder Tabelle sind Beschriftungen, Erläuterungen und ggf. Quellenangaben hinzuzufügen.
- Fußnoten sind erlaubt, sollten aber sparsam verwendet werden.
- Literaturverweise im fortlaufenden Text sowie im Literaturverzeichnis am Ende des Artikels sind nach den Richtlinien des APA7-Stils zu gestalten. Die DOIs der Referenzen müssen hinzugefügt werden, soweit sie über eine verfügen. Zur genauen Zitierweise ziehen Sie bitte das auf den Webseiten der Gesellschaft für Anomalistik bereitgestellte Hinweisblatt heran: https://www.anomalistik.de/zeitschrift/autorenhinweise)
- Autor:innen bleibt räumen dem *Journal of Anomalistics / Zeitschrift für Anomalistik* ein einfaches, nicht-exklusives, weltweites und zeitlich unbefristetes Nutzungsrecht zur Veröffentlichung unter der Creative Commons License Namensnennung-Nicht kommerziell 4.0 (CC BY-NC, https://creativecommons.org/licenses/by-nc/4.0/) ein. Wenn ein Artikel zur Prüfung und Begutachtung beim *JAnom / ZfA* eingereicht wurde, hat die Zeitschrift das Erstveröffentlichungsrecht. Darüber hinaus haben die Herausgeber das Recht, den Artikel im Internet zu veröffentlichen sowie über ein Print-Abonnement verfügbar zu machen.
- Mit der Einreichung versichern die Autor:innen, dass sie über alle erforderlichen Urheber- und Nutzungsrechte am eingereichten Werk verfügen und dass dessen Veröffentlichung keine Rechte Dritter verletzt. Falls der Artikel bereits veröffentlichte Inhalte Dritter enthält (z. B. Abbildungen, Tabellen oder Textauszüge), bestätigen die Autor:innen, dass alle erforderlichen Genehmigungen zur Nutzung und Veröffentlichung vorliegen. Die Autor:innen stellen die Herausgeber und den Verlag von sämtlichen Ansprüchen Dritter frei, die aus einer Verletzung von Urheber- oder Nutzungsrechten im Zusammenhang mit der Veröffentlichung des Artikels entstehen könnten.
- Alle eingereichten Manuskripte unterliegen einer doppelt verblindeten Fachbegutachtung (*double-blind peer review*). Manuskripte, die im Prinzip akzeptabel sind, aber noch nicht den notwendigen Standards entsprechen, werden entweder zur Überarbeitung zurückgeschickt oder redaktionell überarbeitet und dem Autor bzw. der Autorin zur Bestätigung erneut vorgelegt.
- Um die wissenschaftliche Diskussion zu fördern, kann die Redaktion zu einem zur Veröffentlichung angenommenen Manuskript kritische Kommentare und Diskussionsbeiträge von anderen mit dem Thema oder dessen Umfeld befassten Personen anfragen. Autor:innen müssen davon ausgehen, dass solche Beiträge zusammen mit ihrem Artikel publiziert werden. In diesem Fall besteht die Möglichkeit von Autor:innenrepliken.