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 likability 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 $Factor_{10} \geq 10$) 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 individually 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* in *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.

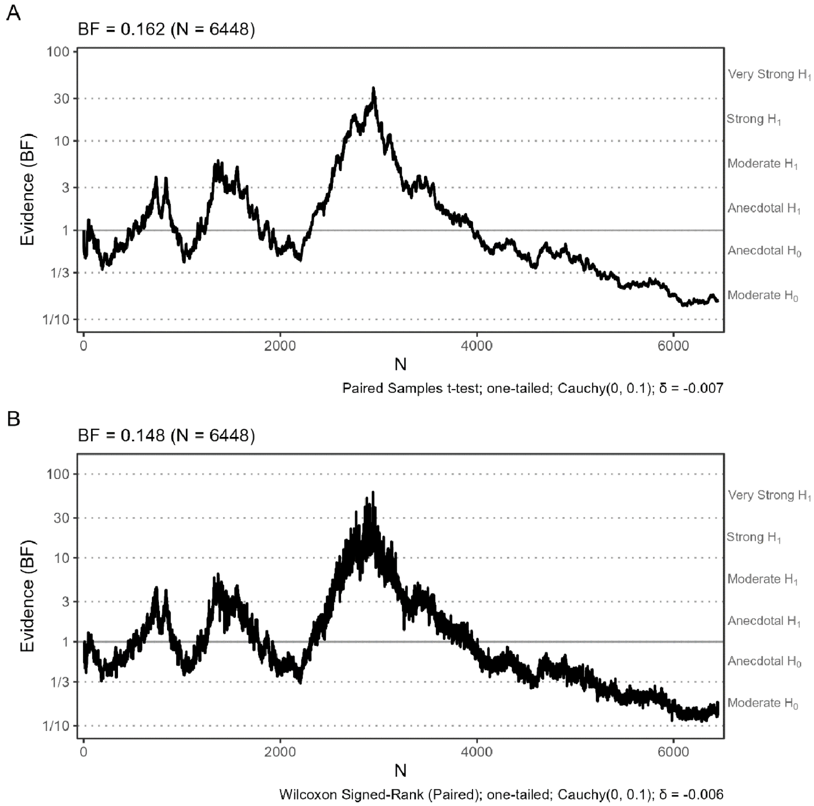
Results

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

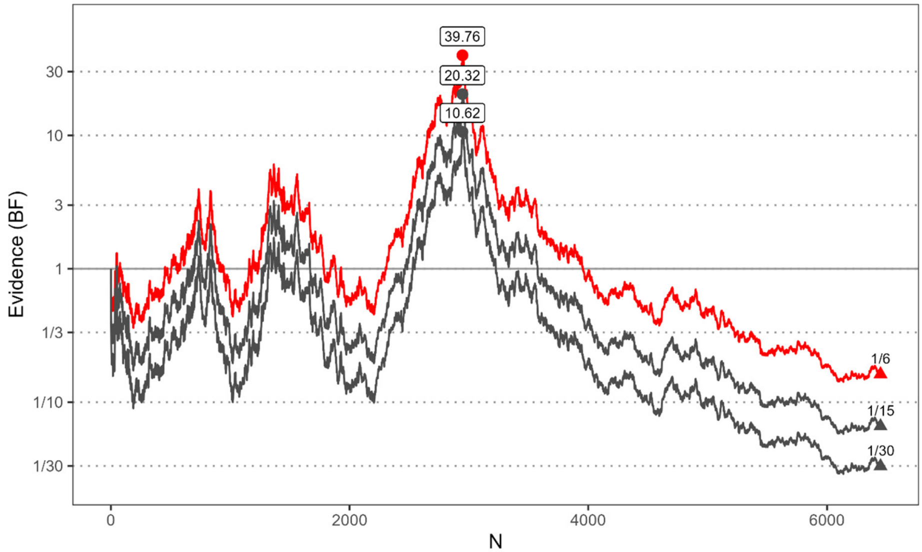
Sequential Bayesian Analyses of the Main Hypothesis Tested With a t-Test (A) and Wilcoxon Signed-Rank Test (B).



In addition, a Bayes factor robustness analysis was performed to explore the impact of the Cauchy prior (δ ~ 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).



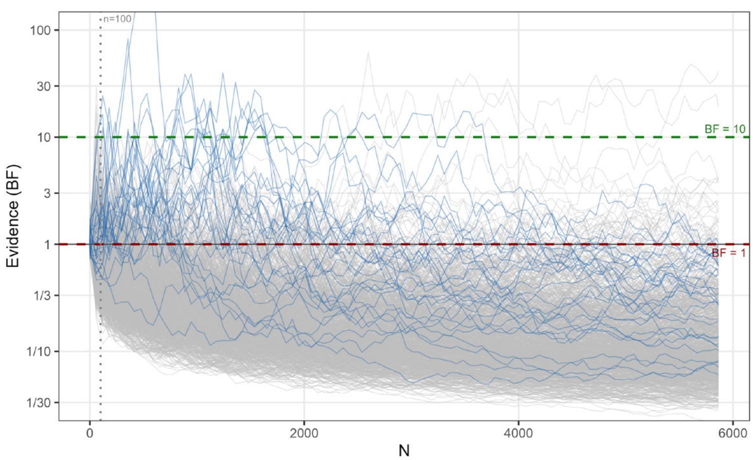
Finally, since the sequential BF of the main analysis showed some unusual fluctuations especially in the first half of the sample which might be caused by extreme outliers in the data, the main analysis was repeated with a non-parametric test (paired Wilcoxon signed-rank test; one-sided). As can be seen 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