Journal of Anomalistics
Zeitschrift für Anomalistik
www.anomalistik.de www.igpp.de

Macroscopic Complementary Relation Between Subjective Observations and Objective Measurements of Colors

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

Abstract – In the Generalized Quantum Theory (GQT), a theory designed to describe psychophysical phenomena, the subjective and objective aspects of reality are considered to be complementary. The research presented here focused on investigating such a macroscopic complementarity between subjective and objective reality descriptions within the context of color assessment. In particular, the non-commutability conjecture of the GQT posits that two observables derived from the subjective and objective subsystems do not simultaneously provide specific eigenvalues. Rather, the act of measurement within one subsystem, e.g. the performance of objective measurements on a stimulus, modifies the state of the entire system, including the eigenvalues of the other subsystem, e.g. the subjective experience of this stimulus. This conjecture was tested empirically in four studies and in three overall analyses of the complete data set combining all four studies. The experimental manipulation involved a measurement variation of one aspect of the supposed complementary pair, namely the storage (non-erasure condition) or deletion (erasure condition) of objective color parameters (hue and lightness). It was anticipated that the erasure manipulation would influence the other part of the pair, namely the subjective evaluations of color with respect to brightness and likability. The primary hypothesis, which tested erasure-dependent effects on variations in subjective brightness scores was confirmed in Studies 1 and 2 but could not be replicated in Studies 3 and 4. The initial findings obtained with this dependent variable can thus be considered false positives and the primary hypothesis could not be confirmed. However, the second hypothesis, which tested erasure-effects on subjective likability mean

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

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

scores revealed the following results: The exploratory findings from Studies 1 and 2, as well as the confirmatory findings from Studies 3 and the preregistered Study 4, indicated an erasure effect on subjective likability, whereby higher likability scores were observed in the non-erasure condition compared to the erasure condition in each study. Since the experimental design was intentionally contaminated by a "color bias" alternative explanation to stabilize the effect documentation, a clear causal erasure interpretation cannot be provided at the single study level. To overcome this contamination an overall analysis strategy was pre-planned and performed on the combined data set. Several overall analyses ruled out the "color bias" alternative explanation and another confound. In conclusion, the overall results provided evidence for an erasure effect on subjective likability supporting the model of GQT. This can be interpreted as macroscopic complementary relations between objective and subjective reality descriptions as proposed by the GQT. The implications of these findings for our understanding of the nature of reality and for the validity of the GQT are discussed.

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

Introduction

The relationship between subjective experience and the external material world has received much attention in philosophy over the millennia (Ruffing, 2021), and psychophysical phenomena have been studied scientifically since the emergence of psychology as an independent academic discipline (Fechner, 1860). The distinction between subjective and objective realities that currently dominates science dates back to René Descartes' substance dualism (Descartes, 1641), which postulates causal relationships between both realities. Two causal relationships between the subjective and objective aspects of our world can be distinguished within dualistic approaches: a) subjective perceptions of the objective world, through which the appearance of objects is translated by our senses into internal experiences about them, and b) volitional processes, through which subjective impulses produce changes in the objective world by force of will. However, it has been argued that such causal interactions are incompatible with substance dualism. Many authors (e.g. Chalmers, 1995; Levine, 1983; Nagel, 1974; Shariff et al., 2008) including Descartes himself (see Perler, 2006) have identified an explanatory gap within this worldview that cannot be easily bridged. Recently, 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; Lucadou et al., 2007; Römer, 2023, 2024; Walach & von Stillfried, 2011) offered an elegant way to overcome these difficulties by proposing macroscopic complementary relations between the subjective and the objective realms. In short, the GQT replaces the causal mechanisms originally thought to underlie this relationship with acausal complementary relations to describe psychophysical interactions. So far, there is no clear scientific evidence for the existence of macroscopic complementary relations (see Römer, 2024). Thus, the main goal of the research presented here was to provide empirical evidence for a macroscopic psychophysical complementary relationship in the area of subjective and objective color assessment.

The GQT was designed to describe complementary phenomena beyond the primary domain of physics, such as psychophysical interactions in psychology, considerations in law, or concepts in art, philosophy, and religious studies, to name a few (Römer, 2024; Walach & Römer, 2011). The central conjectures of GQT were derived from standard quantum theory. GQT systematically adopted some of the core concepts of standard quantum theory while abandoning certain conjectures and restrictions that limited its applications to the physical domain only. Thus, GQT is a theory that describes observations within systems that are not purely physical in nature. The central concepts adopted from quantum theory are complementarity, entanglement and the role of observation (Römer, 2024).

The complementarity principle was originally introduced into quantum physics by Bohr (1928) to explain the dual nature of light, which paradoxically can be described as both a wave and a particle. The two observations are related but complementary, i. e. phenomenologically contradictory, descriptions of the same phenomenon. These contradictory descriptions can be reconciled by considering the particular measurement setup under which either a discrete (particle) or a continuous (wave) state description of light was found (Bohr, 1948, 1949). Another example of a complementarity relation in quantum physics is Heisenberg's (1927) uncertainty relation, which states that position and momentum of a quantum system are complementary features that cannot be simultaneously determined exactly by measurement. In sum, the complementarity principle emphasizes that quantum physical results cannot be defined independently of the measurement setup (see Bohr & Favrholdt, 1999, p. 24–25).

In line with this, Heisenberg (1927, p. 127) emphasized that there is no measurement-independent reality when studying quantum systems. Measuring apparatus and measured system form an inseparable whole in quantum theory. Both form a non-local and acausal entangled state (see Zeh, 1970), and measurement settings are an integral part of the description of quantum states. It is important to note here, that Planck's constant limits this interpretation of reality as measurement-dependent epistemic states to the quantum realm (DeBroglie, 1958). In the macroscopic physical domains covered by classical physics, the measurement dependence can be neglected and reality can be described independently of observational settings. Decoherence theory (Zeh, 1970) confirmed this by showing that, as quantum systems become larger, the critical interaction terms responsible for such complementary relations approach zero through an extension of entanglement, such as in the case of a measurement (Zeh, 2012). Thus, according to standard quantum theory the complementarity principle, entanglement correlations and measurement dependence do not apply to macroscopic physical domains.

In describing phenomena that go beyond purely physical states, the GQT adopts complementarity, entanglement and the role of observation from quantum theory without limiting them to the Planck constant (Römer, 2024). For example, the subjective and objective elements of a psychophysical system constitute subsystems that are considered complementary to each other. Observations made on both subsystems can be incommensurable or incompatible due to their complementary nature (Römer, 2024). 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. In case, the two psychophysical subsystems are spatio-temporally separated, their measurements are supposed to interact non-locally. Although the specific manifestation of such a complementary relation cannot be fully determined theoretically at this stage, at least systematic correlations between the measurement results of the two subsystems are expected when measurement settings are varied (Römer, 2024).

In order to derive an experimental design to test the conjectures of the GQT, the psychophysical system under study, including its potential observables, must be described in more detail. Psychophysical interactions, i.e. transactions between the subjective and the objective realms, involve two classes of phenomena that are well known in psychology: Volition and experience. Wolfgang Pauli and Carl Gustav Jung, in their attempt to unify subjective and objective reality within a dual-aspect monistic approach, provided a terminology for these two classes of psychophysical interactions (for a recent overview, see Atmanspacher & Rickles, 2022). Volition, understood as the subjective, intentional shaping of reality by willpower, resembles in our view an *induced correlation* within in the Pauli-Jung framework (see Atmanspacher & Rickles, 2022). It involves an active, causal effect of an autonomous mind on objective reality. Any specific induced correlation contains a unique volitional impact driven by a certain intention or goal. If stable within an individual or collectively shared by several individuals studied in a sample, this normative impulse is supposed to bias reality formation of an individual or group in line with the shared intention on the participant level. In contrast according to our interpretation of the Pauli-Jung framework, in a *structural correlation* the act of experiencing the world still is described as an active subjective construction process, since subjectivity always involves individual past experiences, preferences, motives and intentions. However, it lacks the normative and stable volitional impulse inherently present in induced correlations. Rather, structural correlations which follow no normative volitional trend usually correlate with a variety of objective physical states (both within an individual across stimuli or in groups of individuals due to the lack of volitional stability). Thus, structural correlations are considered as ad-hoc reality creations following no specific psychological pattern. According to the GQT which proposes acausal, non-local entanglement correlations to describe psychophysical interactions, induced correlations force the acausal correlations into causal local effects even at the participant or sample level (Maier et al., 2022). This would lead to macroscopic time and locality paradoxes which are forbidden in the GQT. The non-transmission (NT) axiom within the GQT was thus formulated to avoid this problematic issue by prohibiting a local causal use of non-local entanglement correlations (Römer, 2024). As a consequence, psychophysical interactions in the form of induced correlations are supposed to be unstable in nature and lead to effect and decline data patterns as described in the model of pragmatic information (MPI; Lucadou et al., 2007). Pauli and Jung intuitively grasped this problem associated with induced correlations and consequently described them as temporarily limited occurring and phenomenologically unstable events (see Atmanspacher & Rickles, 2022). They cannot be objectified by direct replication (Dechamps et al., 2021). In contrast, structural correlations do not refer to normative, stable and/or collectively shared causal relationships between the subjective and the objective reality on the participant or sample level according to our interpretation of the Pauli-Jung framework. Rather, both emerge through individualized and context-specific observations from an originally entangled unified state. Their complimentary relation is a direct consequence of this shared origin establishing an inherently acausal relationship between both. Experience-based psychophysical interactions (= structural correlations) were also considered by Pauli and Jung to be stable and robust in nature on the level of individuals (Atmanspacher & Rickles, 2022).[1]

Recently, the non-commutability conjecture proposed in the GQT, which states that the act of measuring one subsystem changes the state of the whole system, including the eigenvalues of

1 At first sight it may seem that our description of induced and structural correlations somewhat deviates from the original meaning provided by Pauli and Jung (Atmanspacher & Rickles, 2022), since we propose that every subjective perception involves an intentional construction of reality. Rather the core difference between (a) induced and (b) structural correlations according to our definitions is whether a specific psychophysical correlation produces (a) an individual but across several stimuli stable or a stable and across several individuals collective, i. e. normative, reality construction (driven by a specific stable and normative force of will) or (b) a stimulus-specific individualized (i. e. basically random) reality construction lacking this unifying stable and normative force of will. This is in our view the key difference between induced and structural correlations and their impact on reality formation. Consequently, induced correlations due to their subjectively based but normative creation process will run into conflict with an objective reality determined by the laws of nature. They are therefore unstable and cannot be objectified by scientific methods. In contrast, structural correlations just contribute random noise to the objective reality they share what does not stand in conflict with a physicalist description of the world.

the other subsystem, leading to macroscopic complementary relations between subjective and objective observables, has been tested experimentally. In a series of experiments reported by Maier et al. (2024), a psychophysical system based on induced correlations was tested for macroscopic complementary relationships. The system studied consisted of the subjective assessment of one's own intelligence by self-rating, and the objective measurement of that individual's intelligence. The experimental manipulation consisted of an erasure manipulation in which in one condition the objective intelligence data were erased from the result file (overwritten with blank content immediately after each item response) and in another condition these objective intelligence scores were stored. That is, the objective observations or measurements varied between the erased and not-erased conditions. If subjective and objective assessments are incommensurable, as proposed by the GQT, this erasure manipulation should affect the observations of their subjective counterparts. Specifically, it was predicted that subjective intelligence ratings should be higher when objective intelligence scores were erased compared to a non-erased control condition. The psychophysical system tested here was considered an induced correlation, as a collectively shared, normative volitional impulse should be present in subjective intelligence self-ratings, leading to a self-protective overestimation of one's intelligence (Furnham, 2001; see also Gignac & Zajenkowski, 2019; Neubauer & Hofer, 2020). In a pre-registered study (Maier et al., 2024, Study 2), strong evidence for H1 was initially found, i. e., higher subjective intelligence ratings were found in the erasure condition compared to the non-erasure condition, confirming the original hypothesis. However, subsequent replication attempts found evidence for H0 (Bayesian testing approach; Maier et al., 2024, Studies 3 to 7). In a similar approach Maier et al. (2022) tested erasure effects of objective data on subjective reports within a micro-PK task (a typical setting for testing induced correlations characterized by a normative volitional impulse to see favorable stimuli). Similarly, an initial strong evidence for H1, i. e. an erasure-dependent variation of subjective data, was followed by null findings in a replication study (Maier et al., submitted). In summary, when psychophysical systems based on induced correlations were tested for macroscopic complementary relationships between subjective and objective observables, initial evidence was followed by null findings in all subsequent replication attempts. This effect and decline data pattern is consistent with the NT axiom (Lucadou et al., 2007; Römer, 2024), which emphasizes the non-replicability of macroscopic non-local entanglement correlations if they are based on local causal effects characteristic of induced correlations (Römer, 2024). The problem with such effect+decline data is that they appear phenomenologically identical to non-replicable false-positive effects (Dechamps et al., 2021). Thus, a direct replication strategy to test the existence of macroscopic complementary effects derived from induced correlations should be considered a fruitless approach since already at the individual level in addition to the study level causal effects might ultimately destroy the acausal nature of the phenomenon.

The Present Research

In order to convince a skeptical community of the existence of macroscopic complementary relations based on non-local entanglement correlations within psychophysical systems our previous focus on induced correlations (Maier et al., 2022, 2024) was abandoned (since they inherently refer to classical causality both at the participant level due to their normative intention and – when tested scientifically – at the experimenter level) in favor of structural correlations. Structural correlations were considered as robust and replicable at least on the level of participants by Pauli and Jung (see Atmanspacher & Rickles, 2022). They are defined as non-normatively appearing psychophysical correlations that do not require a collective direct causal relationship mediated by the *unus mundus* when individual experiences within a heterogenous sample are considered. Rather, subjective experiences and their objective counterparts evolve acausally from a common underlying state during the act of experiencing. According to our interpretation of the GQT, these psychophysical systems should be stable when scientifically documented under appropriate conditions (see below), and systematic macroscopic complementary relations should then be found when measuring the respective subjective and objective observables.

The term "appropriate conditions" refers to the problem of applying causality tests when investigating acausal phenomena. Psychophysical complementary relations in the form of structural correlations are characterized by acausal interrelations between the objective and subjective realms at the individual level. Thus, the original structural correlation when applying an experimental causality test might take on features of an induced correlation on the experimenter level leading to effect and decline data pattern as reported above. Consequently, Römer (2024, p. 56), based on the conjectures derived from the GQT, emphasized that in case such interrelations are tested in laboratory experiments one should not focus on isolated causal influences of experimental manipulations. Rather unusual correlations between the two subsystems (e.g. the objective and subjective) in combination with the experimental manipulation should be described empirically. That is, the experimental causality test needs to be weakened to some extent in order to find a potentially robust effect. We took this recommendation serious and this aspect was most important for the development of an appropriate experimental design in our present research. In line with Römer (2024), we decided to attenuate the direct causality test within the design by reducing each study's internal validity. We tried to achieve this goal by applying the following measures: First, an experimental manipulation of objective measurements of stimuli was invented (erasure-manipulation) that should causally impact their complementary counterparts, the subjective assessments of these stimuli (= classical scientific method). Second, an additional variable was introduced that could alternatively explain the results. In our case, small subsets of stimuli were randomly pre-selected and randomly assigned to the control and experimental condition for subjective evaluation. This leaves room for the alternative explanation that when differences in subjective evaluations between the experimental conditions indicating evidence for H1 were found in a study, they could have alternatively been caused by a subjective assessment bias resulting from the random assignments of only limited subsets of stimuli to these experimental conditions. That is, with only a few colors used in each experimental condition, differences found favoring H1 can alternatively be attributed to chance.

In this way, the exact causal nature will remain obscure when each study is considered separately, hopefully allowing potentially robust macroscopic complementary relations to be found. Finally, across several studies, the alternative factor (biased stimuli selection favoring H1 due to chance) will then be ruled out retrospectively in final overall analyses of the effects found by testing the role chance might have played in the stimuli subsets' assignment to experimental conditions. If successful, these overall analyses will rectify the internal validity violation within the experiments. This would allow to satisfactorily and retrospectively interpret the effects of the experimental manipulation on our DVs. This research strategy clearly deviates from standard scientific approaches in which causal effects of independent variables on dependent variables are to be documented in isolation without confounding alternative variables. However, this approach seemed to be the most promising given the acausal nature of the phenomenon under study.

In the studies presented here, subjective experiences of color were assessed under variations of measurements of objective color parameters. Subjective color experience was assumed to be a rather non-normative perception process. Although color preferences exist (e.g. Palmer & Schloss, 2010) their assessment should not be biased normative-intentionally to create a specific collective experience. This is in contrast to intelligence self-assessments, which involve active and normative intentional constructions of reality (Furnham, 2001). Thus, it was considered an appropriate candidate for establishing structural correlations as defined here. As independent variable, the objective color parameters *hue* and *lightness* of randomly selected colors were either stored in a result file or omitted from the result file and permanently erased after completion of the study (erasure manipulation). In addition, only limited subsets of color stimuli were selected for each experimental condition to provide an alternative explanation of the effects (biased colors favoring H1 due to chance). Subjective brightness ratings and subjective liking of the colors served as dependent variables. In four studies, erasure effects of objective color parameters on variations in subjective brightness ratings and on mean score differences in subjective likability ratings were tested.

If (a) subjective experiences vary with the erasure manipulation consistently in each study (Step 1), if (b) such findings are replicable (Step 2) and if (c) the alternative factor (the biased color argument) can be ruled out retrospectively by final overall analyses (Step 3), the existence of macroscopic complementary relations as proposed by the GQT would be confirmed. This three-step research agenda was applied to the studies described here. Note, that if the term “erasure-dependent effect”, “erasure manipulation effect” or similar expressions are used in the following sections throughout the studies’ description it also includes the potential alternative explanation of a “biased color effect”. The use of these expressions therefore does not presuppose a specific interpretation of the experimental manipulation effect at least until the final results of the overall analyses have been reported.

We decided to investigate erasure-dependent effects on *variations* in subjective brightness ratings rather than on mean differences between subjective brightness ratings themselves, which was based on an analogy to the Heisenberg uncertainty relation (Heisenberg, 1927). According to this approach, certainty (non-erasure condition) or uncertainty (erasure condition) about objective color parameters should influence the degree of subjective (un)certainty, operationalized as individual variations of brightness ratings within the experimental conditions. Filk and Römer (2011) referred to such potentially existing macroscopic complementarity relations as strong complementarity. However, they emphasized that other forms of macroscopic complementary relations could also exist according to the GQT. Therefore, erasure-dependent mean score differences in subjective liking were also tested.

In Studies 1, 2, and 3, erasure-dependent variations in brightness ratings were tested in a confirmatory manner and in Studies 1 and 2, erasure-dependent subjective likability mean score differences were tested in an exploratory manner. After the results of the first two studies were known, erasure-dependent differences in subjective likability mean scores were tested in a confirmatory manner in Study 3 and in a pre-registered Study 4. The exact hypotheses (confirmatory or exploratory) can be found at the beginning of the results sections of each study. In addition to these first two steps (when successful), in Step 3 overall analyses including color subset permutations (a priori) and multi-level models (MLM; post-hoc) will be provided at the end of the studies description section to address the potential impact of the alternative “biased color subsets favoring H1 by chance” explanation.

Study 1

Methods

Ethical Guidelines

At the beginning of the online experiment, participants were given general information about the study. The voluntary nature of their participation and the privacy policy were emphasized. Participants could accept participation by pressing a button (informed consent procedure). All stored data were encrypted and analyzed anonymously. Ethical approval for the study was obtained from an ethics committee.

Sample

Bayesian sequential analysis procedures were used to analyze the data. This approach permitted the cumulative collection and analysis of data, allowing for the testing of additional participants and the successive addition of new data to the dataset until a specific Bayes Factor (BF) for H1 or H0 was reached. A priori, a minimum evidential criterion of BF ≥ 10 (strong evidence) was set. This entailed monitoring the data and continuing data collection until a BF ≥ 10 in favor of H0 or H1 was reached. Should this occur, data collection was terminated and the final BFs were reported. The minimum n to start Bayesian testing in this and all subsequent studies was set to $n = 100$.

The participants were recruited primarily through private contacts and social media by LMU students in the context of experimental courses and bachelor's theses (see acknowledgements) 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.

A total of 974 participants took part in the study. As defined a priori, participants were excluded from the data analyses if they indicated that their responses were not reliable ("data integrity item," $n = 8$) or if they indicated that they were color blind ($n = 21$). The final sample included in the analyses consisted of $N = 945$ participants. The study was declared complete at this N since BF10s ≥ 10 were reached in the Bayesian analyses performed.

The sample consisted of 535 female, 395 male, and 13 diverse participants. The mean age of the participants was 29.95 years ($SD = 14.71$; max. age: 79). Two individuals did not provide demographic information. Data collection was conducted online via PC, tablet, or smartphone, and participants completed the study in German or English language.

Materials

Colors Used in the Experiment: Objective Color Parameters

In advance of the experiment, six distinct colors were randomly selected and assigned to either the erasure or non-erasure condition prior to data collection, serving as experimental stimuli in this study. Three standard color parameters from the HSL model (H = hue, S = saturation, L = lightness) were employed to define a selected color. The first parameter was hue which spans a range from 0 to 359 degrees on a standard color circle. A QRNG was employed to randomly determine the hue of each of the six colors by selecting a random number out of the aforementioned range. The second parameter was saturation which typically ranges from 0% to 100%. This parameter was set at 100% and maintained at that level for each color utilized in this study. The third parameter was lightness, which indicates the brightness of a color and typically ranges from 0% to 100%. In the non-erasure condition, the lightness scores of 60%, 50%, and 40% were selected a priori and randomly assigned to the three control colors. This should ensure a relatively dispersed lightness score distribution within this limited set of three colors. The lightness scores for the experimental color stimuli selected for use in the erasure condition were randomly selected from a restricted lightness scale ranging from 30% to 70%. This was done to achieve a random but potentially similar lightness score distribution within this small set of three colors, as in the non-erasure control condition. Additionally, this process allows the hue of the colors to remain recognizable. The random parameter selection process for the erasure condition was conducted via a Python script, which stored the parameters in a separate file, thus obviating the necessity for a researcher to inspect them. Consequently, the objective color parameters were not subjected to inspection during the selection process nor while data collection was being performed.

The objective color parameters of the non-erasure condition were stored directly in the experimental file and were accessible to the researchers for subsequent inspection after data collection. The actual color parameters were as follows: hsl(164, 100%, 40%), hsl(199, 100%, 50%) and hsl(38, 100%, 60%). The objective color parameters of the three colors utilized in the erasure condition were temporarily stored in a separate file during the experiment but not stored permanently. This file containing the data was deleted following the conclusion of data collection without undergoing any form of inspection to ascertain that the parameters could never be identified after the experiment. 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, 2024) and non-local phenomena should also occur in a time-reversed manner as long as the erased information has not been inspected at any time before being destroyed, which was the case.

The hue of the erasure condition colors exhibited a random variation between 0 to 359, while the preset saturation was maintained at 100% for all colors and the lightness ranged between 30% and 70%. Therefore, an uncertainty with regard to objective hue and objective lightness was established for the experimental color stimuli. The experimental program and color selection scripts for this and all other studies can be retrieved from OSF (https://osf.io/u2zqp).

During data collection each participant was randomly assigned to one of the six colors. This color was presented as a rectangle (400 by 400 pixels; size dependent on screen used) in the center of the screen for subjective color assessments.

The use of limited subsets of colors (3 x 3) randomly assigned to each experimental condition established an alternative explanation for the potential effects to be found in this study, since colors favoring the confirmation of H1 can be selected in the respective condition by chance. Although this aspect of the design undermines a clear causal interpretation of the effect potentially found with the erasure manipulation, it was intentionally established to stabilize the acausal nature of the complementary relation tested in this and all subsequent studies. As mentioned above, this alternative explanation will be addressed in final overall analyses at the end of the studies section in case Steps 1 and 2 were completed successfully.

Subjective Color Assessment

The color presented to each participant was subjectively rated on two characteristics: brightness and likability. The color rectangle was initially evaluated based on its subjective brightness. This was assessed by means of the question presented beneath the color: "How bright is the color?" (German original: "Wie hell ist die Farbe?"). A visual analogue scale, comprising a grey horizontal bar, was provided beneath the question, ranging from "extremely dark" (0) to "extremely bright" (100). Only the labels at the end of the scale were visible, not any numerical values. Participants were instructed to move a slider to the left or the right, which initially appeared at the midpoint of the scale. Following the assessment of brightness, the color rectangle remained on the screen and the subjective likability of the color was measured by the question presented beneath the color: "How much do you like the color?" (German original: "Wie gut gefällt Ihnen die Farbe"?). This question was combined with the same 101-point visual analogue scale as described above ranging from "not at all" (0) to "very much" (100).

Procedure

Participants were provided with a general overview of the study procedures. Subsequently, the participants provided informed consent to participate in the study. A pseudorandom number generator (jsPsych's *randomization* module) was employed to determine the color to be presented to the participant and, consequently, their assignment to either the erasure or non-erasure condition. This module relies on JavaScript's Math.random() function to generate pseudorandom numbers. A pseudorandom number generator is an algorithm that produces a sequence of numbers that mimic the properties of random numbers while being generated deterministically from an initial seed value. This method is commonly employed in experimental psychology to ensure the integrity and validity of research findings.

It is crucial to highlight that in this and all subsequent experiments, both the participants and the experimenters were unaware of any objective data erasure or storage manipulations. Furthermore, participants were never informed about their assigned condition or the existence of these conditions.

Following this, the color rectangle was displayed in the center of the screen along with the first inquiry regarding the perceived brightness of the color and the visual analogue scale, as previously described. Subjective evaluations of the brightness of the color were requested ("How bright is the color?"), with participants moving the slider to the position on the scale that best reflected their assessment of the color. The final rating was then confirmed by pressing the appropriate button. Next, the color rectangle was still present but now accompanied by the likability question, "How much do you like this color?" together with the corresponding visual analogue scale. Participants were asked to subjectively evaluate the likability of the color by moving the slider to the appropriate position between the extremes of the scale. The final rating was then confirmed by selecting the appropriate button.

Following the completion of the ratings, participants were asked to provide their age and gender. Finally, two items were presented, asking whether the participants wished for their responses to be included in the analysis. This item was entitled "Hand on your heart: Did you really work on this study attentively and conscientiously so that we can use your data?" and participants were invited to respond with either "yes" or "no". Additionally, participants were asked whether they were colorblind, with the options "yes" and "no" provided.

Subsequently, the results were then stored in a result file which contained either solely subjective color data (in the erasure condition) or subjective and objective color data (in the non-erasure condition), depending on the condition.

Design and Statistical Analysis

The study design was a between-subjects design with an independent variable consisting of two conditions: the availability (non-erasure condition) or non-availability (erasure condition) of the objective color parameter scores in the final results file.

The primary statistical analysis entailed a two-tailed Bayesian independent samples t-test with two independent groups (non-erasure vs. erasure) serving as the independent variable (IV) and the variation mean score of subjective brightness ratings as the dependent variable (DV). The DV was calculated as the absolute difference between each individual brightness rating score and the mean brightness score for the respective color group. For each of the six colors utilized in this study a group mean score of brightness ratings was calculated for that purpose. Subsequently, the absolute differences were averaged to obtain an overall brightness variation mean score for each color. Finally, the three variation mean scores of the three colors in the erasure condition were averaged to obtain one overall mean score. The same procedure was followed with the three variation mean scores obtained from the three colors in the non-erasure condition. This overall variation mean score, which described the variability of subjective brightness ratings within each condition, served as DV in the main analysis.

In addition to the primary analysis, an analogous exploratory analysis was conducted using subjective likability rating mean score as the DV. This analysis employed a two-tailed Bayesian independent samples t-test with two independent groups (non-erasure vs. erasure) as the independent variable (IV) and the mean score of subjective likability ratings as the dependent variable (DV). The analysis script for this and all other studies can be found at OSF (https://osf.io/u2zqp).

Parametric t-testing was employed in this and all subsequent studies, as the equality of variances assumption was met in all cases (visual inspection approach). The normality assumption was not subjected to systematic scrutiny, given that independent t-tests have been demonstrated to remain robust against violations of normality when the sample size is large ($N > 500$; Lumley et al., 2002).

In both Bayesian t-tests, an a priori uninformed prior was employed, based on an estimated effect size of $d = .1$, following a Cauchy distribution centered around 0 with $r = .1$ (i. e., $\delta \sim$ Cauchy [0, .1]). Data collection was conducted sequentially, with data accumulated in the order of participation, and analyzed in chronological order, based on the date of completion of the ratings by the participants.

Results

Main Analysis

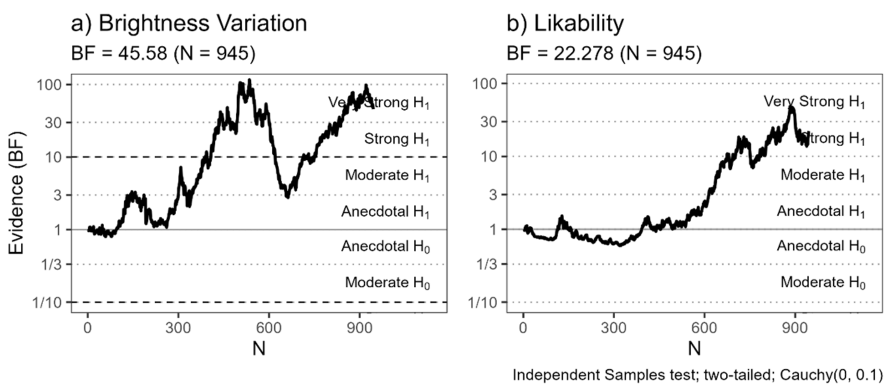
To test the primary hypothesis that an erasure manipulation of the objective color parameters data (erasure vs. non-erasure) affects the mean variation of subjective brightness ratings (undirected hypothesis), a two-sided independent samples Bayesian t-test was conducted with experimental condition (erasure vs. non-erasure) as IV and mean variation of subjective brightness ratings as DV. The Bayesian independent sample t-test (two-tailed, $N = 945$) yielded a final $BF_{10} = 45.58$, indicating very strong evidence in support of H1 ($d_{cohen} = .23$). The BF first reached the threshold of BF ≥ 10 at $n = 389$ (see Figure 1a). The mean variation of brightness score was higher in the non-erasure group ($M = 12.45$, $SD = 9.22$, $n = 454$) than in the erasure group ($M = 10.40$, $SD = 8.72$, $n = 491$).

Exploratory Analyses

In addition to the primary analyses presented above, an exploratory analysis was conducted to investigate the impact of a color parameter erasure effect on the mean likability rating score. A two-tailed Bayesian independent samples t-test was performed, with erasure manipulation (erasure vs. non-erasure) as the IV and mean likability rating as the DV. The independent samples Bayesian t-test (two-tailed, $N = 945$) yielded a final $BF_{10} = 22.28$, indicating strong evidence in support of H1 ($d_{cohen} = .21$; see Figure 1b). The mean likability score was higher in the non-erasure group ($M = 58.56$, $SD = 24.86$, $n = 454$) than in the erasure group ($M = 52.87$, $SD = 28.53$, $n = 491$).

Figure 1

Accumulation of Bayesian Evidence for the Confirmatory Hypothesis (a) and the Exploratory Hypothesis (b) of Study 1



Discussion

The primary objective of Study 1 was to find evidence for erasure-dependent objective color parameter effects on subjective color ratings. Specifically, in the primary analysis, we examined the impact of objective color parameter uncertainty on subjective brightness ratings. We tested whether an experimental manipulation of objective color parameter uncertainty, including lightness (erasure condition = complete uncertainty vs. non-erasure condition = complete certainty), affected the uncertainty of subjective brightness ratings (high vs. low variations in subjective brightness ratings). A Bayesian independent samples t-test with experimental condition (erasure vs. non-erasure of objective color parameters) as IV and mean variation of subjective brightness ratings as DV indicated very strong evidence for H1. The results demonstrated a higher degree of variation in subjective brightness in the non-erasure condition compared to the erasure condition. Furthermore, the exploratory analysis indicated a comparable, though unanticipated, experimental impact of the IV on subjective likability. A Bayesian independent samples t-test with experimental condition (erasure vs. non-erasure of objective color parameters) as IV and mean likability rating as DV also indicated strong evidence for H1. The results indicated that participants rated the color presented in the non-erasure condition more favorably than in the erasure condition. The effect sizes obtained in both analyses indicated small effects of the erasure manipulation.

It is important to note that the results observed in Study 1 could also be attributed to the random selection process employed in the choice of colors utilized as experimental stimuli within each subset. It is possible that the colors selected to constitute a subset were randomly prone to specific subjective brightness variations or specific likability scores, resulting in a greater number of "variations in brightness inducing" and "more likeable" colors in the non-erasure condition compared to the erasure condition. This alternative explanation was included 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 conjecture derived from the GQT (Atmanspacher et al., 2002; Walach & Römer, 2000, 2011; see also Fach, 2011; Filk & Römer, 2011; Hinterberger & von Stillfried, 2013; Lucadou et al., 2007; Römer, 2023, 2024; Walach & von Stillfried, 2011) that macroscopic complementarity relations might exist between objective and subjective measurements of color features. These effects would then highlight a central feature of the GQT, namely the non-commutability conjecture (Römer, 2024). This conjecture predicts that the existence or non-existence of objective measurements about colors affects their subjective evaluation, both with regard to uncertainty of brightness and likability of the colors. Or, since there were only few colors to be assessed, the data can be explained by chance. This alternative explanation is currently the more probable and must be disproven by subsequent independent replication attempts and an overall analysis addressing the likelihood of occurrence for such biases by chance. Only if the effects found in Study 1 (Step 1) are consistent across different random color selections within different experiments (Step 2) and overall analyses can exclude the biased color argument (Step 3) the true nature of the effect can be determined. Consequently, Study 2 was conducted to replicate Study 1.

Study 2

The aim of Study 2 was to replicate the effects observed in Study 1. The replication was nearly identical, with the exception of six new randomly selected colors and a single minor alteration to the control color stimuli selection, which will be elaborated upon subsequently. The primary objective was to replicate the experimental manipulation effect identified in the primary analysis, with the subjective brightness variation mean score designated as the DV. Additionally, an exploratory investigation was conducted to ascertain whether a comparable experimental manipulation effect would be observed with the likability mean score as the DV. At this juncture, there was a question as to whether the exploratory finding from Study 1 represented a genuine effect. Consequently, no strong expectations were held regarding this DV in Study 2, and it was tested again in an exploratory manner.

Methods

Ethical Guidelines

At the beginning of the online experiment, participants were informed about the study in general terms. The voluntary nature of their participation and the data protection regulations were emphasized. Participants could accept participation by clicking a button (informed consent procedure). All stored data were encrypted and analyzed anonymously. Ethical approval for the study was obtained from an ethics committee.

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 ≥ 10 (strong evidence) serving as stopping rule in the main analysis. If this was the case, data collection was terminated and the final BFs were subsequently reported.

The participants were recruited primarily through private contacts and social media by LMU students in the context of experimental courses and bachelor's theses (see acknowledgements) 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.

A total of 773 participants participated in the study. As defined a priori, participants were excluded from the data analyses if they indicated that their responses were not reliable ("data integrity item," $n = 16$) or if they indicated that they were colorblind ($n = 10$). The final sample included in the analyses consisted of $N = 749$ participants. The study was declared complete at this N since BF10s ≥ 10 were reached in the Bayesian analyses performed.

The sample consisted of 488 female, 251 male, and 7 diverse participants. The mean age of the participants was 32.63 years ($SD = 15.97$; max. age: 100). Three individuals did not provide demographic information. Data collection was conducted online via PC, tablet, or smartphone, and participants completed the study in German or English language.

Materials

Colors Used in the Experiment: Objective Color Parameters

In advance of the experiment, six distinct colors were randomly selected and assigned to either the erasure or non-erasure condition prior to data collection, serving as experimental stimuli in this study. The random color selection process followed the same protocol as in Study 1 with the following single exception: In the non-erasure condition, the three control colors were selected based on their preselected lightness scores of 65%, 50%, and 35%, which were randomly assigned to the three control colors. This constituted the sole discrepancy between the color selection processes employed in Study 1 and Study 2. The decision to utilize more extreme lightness scores for the control colors as in Study 1, was predicated on the objective of ensuring a more dispersed lightness score distribution encompassing the entire lightness spectrum more accurately within this limited set of three control colors in the non-erasure condition of Study 2, as compared to Study 1. The lightness scores for the experimental color stimuli selected for use in the erasure condition were once again randomly selected from a restricted lightness scale ranging from 30 to 70. This was done to achieve a random but potentially similar lightness score distribution within this small set of three colors as in the non-erasure control condition. The objective color parameters were not examined during this process or while data collection was being conducted.

The objective color parameters of the non-erasure condition were stored in a result file and subsequently inspected. The actual color parameters were as follows: hsl(178, 100%, 65%), hsl(242, 100%, 35%), and hsl(316, 100%, 50%). The objective color parameters of the three colors utilized in the erasure condition were not stored in the result file; their information was subsequently deleted upon completion of data collection to ensure that they could never be identified. The hue of the erasure condition colors exhibited random variation between 0 and 359, while the preset saturation was 100% for all colors and the lightness varied between 30% and 70%. This resulted in an uncertainty regarding the objective hue and objective lightness of the experimental color stimuli.

During data collection each participant was randomly assigned to one of the six colors. This color was presented as a rectangle (400 by 400 pixels; size dependent on screen used) in the center of the screen for subjective color assessments.

Subjective Color Assessment

The color presented to each participant was subjectively rated on two characteristics: brightness and likability. These ratings were made on visual analogue scales ranging from 0 to 100 as was done in Study 1 (for details see corresponding section of Study 1).

Procedure, Design and Statistical analyses

The procedure, design, and statistical analyses including the Cauchy prior were identical to those employed in Study 1.

Results

Main analysis

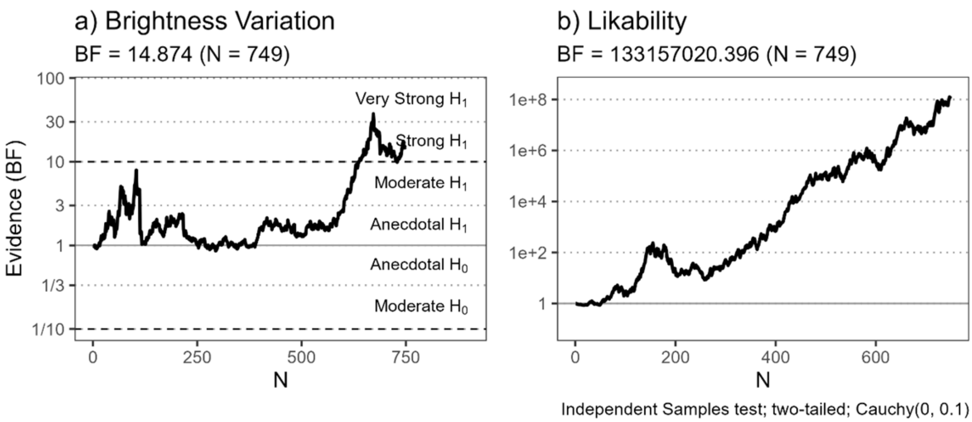
To test the primary hypothesis that an erasure manipulation of the objective color parameters data (erasure vs. non-erasure) affects the mean variation of subjective brightness ratings (undirected hypothesis), a two-sided independent samples Bayesian t-test was conducted with experimental condition (erasure vs. non-erasure) as IV and mean variation of subjective brightness ratings as DV. The Bayesian independent sample t-test (two-tailed, $N = 749$) yielded a final $BF_{10} = 14.87$, indicating strong evidence in support of H1 ($d_{cohen} = .23$). The BF first reached the threshold of $BF \geq 10$ at $n = 638$ (see Figure 2a). The mean variation of brightness score was higher in the non-erasure group ($M = 12.76$, $SD = 9.75$, $n = 382$) than in the erasure group ($M = 10.57$, $SD = 9.47$, $n = 367$).

Exploratory Analyses

In addition to the main analysis, an exploratory analysis was conducted testing a color parameter erasure effect on the mean likability rating score. A two-tailed Bayesian independent samples t-test with erasure manipulation (erasure vs. non-erasure) as the IV and mean likability rating as the DV was performed. This independent samples Bayesian t-test (two-tailed, *N* = 749) yielded a final BF_{10} = 133.2 million, indicating extreme evidence in support of H1 (d_{cohen} = .50; see Figure 2b). The mean likability score was higher in the non-erasure group (*M* = 56.87, *SD* = 27.58, *n* = 382) than in the erasure group (*M* = 43.27, *SD* = 27.18, *n* = 367).

Figure 2

Accumulation of Bayesian Evidence for the Confirmatory Hypothesis (a) and the Exploratory Hypothesis (b) of Study 2



Discussion

The objective of Study 2 was to replicate the findings of Study 1 with a distinct set of color stimuli. Specifically, in our primary analysis, we once again concentrated on the brightness of colors and tested whether an experimental manipulation affected the uncertainty of subjective brightness ratings (high vs. low variations in subjective brightness ratings). A Bayesian independent samples t-test with the experimental condition (erasure vs. non-erasure of objective color parameters) as IV and the mean variation of subjective brightness ratings as DV revealed strong evidence for H1. The results demonstrated a higher variation of subjective brightness ratings in the non-erasure condition compared to the erasure condition, which successfully replicated the effects obtained with the same DV in Study 1.

The exploratory analysis also yielded an experimental manipulation effect on subjective likability. A Bayesian independent samples t-test with experimental condition (erasure vs. non-erasure of objective color parameters) as IV and mean likability rating as DV revealed extreme evidence for H1. Participants indicated a greater liking for the color presented in the non-erasure condition than in the erasure condition, which replicated the same effect found in Study 1.

The effect sizes obtained in both analyses indicated small to medium effects of the experimental manipulation. Again, a clear interpretation of the experimental manipulation effect on subjective ratings cannot be provided here, since the effect can either be caused by the erasure manipulation itself or by a biased random color subset assignment favoring the H1 by chance due to the limited subsets of three colors in each experimental condition, or both.

It is important to note that our findings are also limited in terms of the used design. This limitation should be acknowledged and addressed in future replication attempts. The random color selection process was conducted in a manner that differed between the erasure and non-erasure conditions in both Study 1 and Study 2. While the colors for the erasure manipulation were randomly selected with regard to hue (from 0 to 359) and lightness (from 30% to 70%), the colors selected in the non-erasure condition were only randomly selected for hue but preset for lightness (40%, 50%, 60% in Study 1 and 35%, 50% and 65% in Study 2, respectively). Although the decision to proceed with this procedure was justified by compelling reasons, namely to guarantee an equal distribution of lightness scores across the complete scale utilized in the experiments and to achieve a perfect control color subset, the disparate treatment of color selection procedures for each color subset employed in the two experimental conditions introduced an additional confounding factor that could potentially account for the observed results. To address these limitations, the following replication attempt excluded the lightness-confound by applying the same color selection procedures to both experimental conditions. It also further minimized the "biased colors" argument by increasing the randomly selected colors for each subset from three to six in each experimental condition.

Study 3

The objective of Study 3 was to conceptually replicate the effects obtained in Studies 1 and 2. The experimental design was largely similar to that of the preceding studies with the following exceptions: (a) in the present study 12, different colors were randomly selected as experimental stimuli, (b) the random selection procedure for all colors now followed the exact protocol for the color selection procedure of the erasure color subset in Study 1 and 2. In other words, the same random selection procedure was applied in both experimental conditions, effectively ruling out the aforementioned lightness-related confounding factor, and (c) both effects documented thus far were tested in a confirmatory manner.

The objective of Study 3 was to conceptually replicate the experimental manipulation effects observed in the main analyses with the subjective brightness variation mean score as DV. Additionally, the study aimed to conceptually replicate the experimental manipulation effects observed with the likability mean score as DV. We decided to increase the number of colors to gradually decrease the likelihood of the alternative 'color bias' explanation.

Methods

Ethical Guidelines

At the beginning of the online experiment, participants were informed about the study in general terms. The voluntary nature of their participation and the data protection regulations were emphasized. Participants could accept participation by clicking a button (informed consent procedure). All stored data were encrypted and analyzed anonymously. Ethical approval for the study was obtained from an ethics committee.

Sample

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

Participants were recruited through several bachelor's and master's theses (see acknowledgements) 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.

A total of 584 participants took part in the study. As defined a priori, participants were excluded from the data analyses if they indicated that their responses were not reliable ("data integrity item," n = 12) or if they indicated that they were color blind (n = 11). The final sample included in the analyses consisted of N = 562 participants. The study was declared complete at this N since at least one of the BF $\geq$ 10 was reached in the Bayesian analyses performed.

The sample consisted of 362 female, 198 male, and 2 diverse participants. The mean age of the participants was 30.01 years (*SD* = 13.35; max. age: 83). Data collection was conducted online via PC, tablet, or smartphone, and participants completed the study in German language.

Materials

Colors Used in the Experiment: Objective Color Parameters

In advance of the experiment, 12 new colors were randomly selected and designated as experimental stimuli for the purposes of this study. The random color selection process followed the same protocol as in Study 1 and 2 with the following single exception: In the non-erasure condition, the lightness scores of the six control colors were randomly selected from a lightness range between 30% and 70% by QRNG. This was the sole distinction between the color selection processes employed in Studies 1 & 2 and Study 3. Consequently, both color subsets were randomly selected with regard to hue (0 to 359) and lightness (30% to 70%). Once more, saturation was preset at 100% and remained constant for all 12 colors. The objective color parameters were neither examined during this process nor while data collection was conducted.

The objective color parameters of the non-erasure condition subset were stored in a result file and subsequently analyzed. The actual color parameters were as follows: hsl(317, 100%, 32%), hsl(164, 100%, 51%), hsl(350, 100%, 64%), hsl(85, 100%, 54%), hsl(19%, 100%, 57%), and hsl(203, 100%, 47%). The objective color parameters of the six colors utilized in the erasure condition were not stored, and the file containing the data was deleted upon completion of the data collection process without undergoing any form of inspection to ensure that the data could never be identified. The hue of the erasure condition colors exhibited random variation between 0 and 359, while the preset saturation was 100% for all colors. The lightness of these colors demonstrated variability between 30% and 70%. Therefore, an uncertainty with regard to objective hue and objective lightness was established for the experimental color stimuli as a result of this process.

During data collection each participant was randomly assigned to one of the 12 colors. This color was presented as a rectangle (400 by 400 pixels; size dependent on screen used) in the center of the screen for subjective color assessments.

Subjective Color Assessment

The color presented to each participant was subjectively rated on two characteristics: brightness and likability. These ratings were made on visual analogue scales ranging from 0 to 100 as was done in Studies 1 and 2 (for details see corresponding section of Study 1).

Procedure, Design and Statistical analyses

The methodology employed in this study was identical to that utilized in Studies 1 and 2, encompassing the same procedures, design, and statistical analyses, including the Cauchy prior.

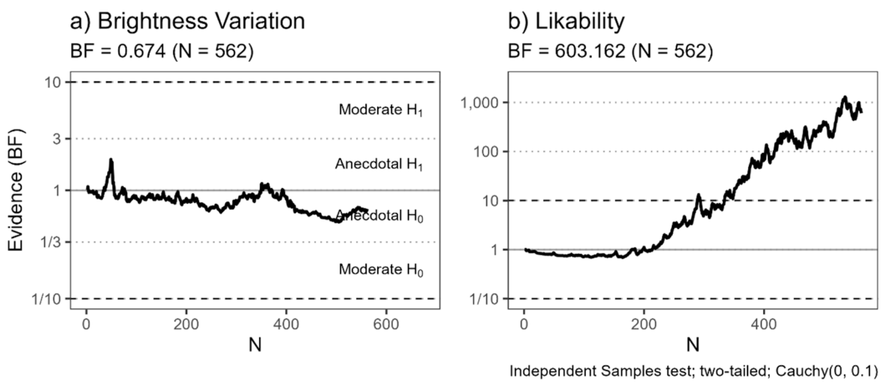
Results

To test the first hypothesis that the experimental manipulation (erasure vs. non-erasure) affects the mean variation of subjective brightness ratings (undirected hypothesis), a two-sided independent samples Bayesian t-test was conducted with experimental condition (erasure vs. non-erasure) as IV and mean variation of subjective brightness ratings as DV. The Bayesian independent sample t-test (two-tailed, $N = 562$) yielded a final $BF_{01} = 1.63$, indicating anecdotal evidence in support of H0 ($d_{cohen} = -.09$; see Figure 3a). The mean variation of the brightness score was found to be lower in the non-erasure group ($M = 10.83$, $SD = 8.34$, $n = 272$) than in the erasure group ($M = 11.68$, $SD = 10.14$, $n = 290$).

To test the second hypothesis, a further analysis was conducted to examine the impact of the experimental manipulation on the mean likability rating score. A two-tailed Bayesian independent samples t-test was conducted with experimental manipulation (erasure vs. non-erasure) as the IV and mean likability rating as the DV. The independent samples Bayesian t-test (two-tailed, $N = 562$) yielded a final $BF_{10} = 603.16$, indicating extreme evidence in support of H1 ($d_{cohen} = .37$). The BF first reached the threshold of $BF \geq 10$ at $n = 289$ (see Figure 3b). The mean likability score was higher in the non-erasure group ($M = 60.47$, $SD = 26.21$, $n = 272$) than in the erasure group ($M = 50.86$, $SD = 26.18$, $n = 290$).

Figure 3

Accumulation of Bayesian Evidence for Both Confirmatory Hypotheses (a) and (b) of Study 3



Discussion

The objective of Study 3 was to conceptually replicate the findings of Studies 1 and 2 using a different and extended set of color stimuli (12 colors instead of 6) and a complete random color selection procedure for both conditions to eliminate the potential for lightness related confounding variables. In the first analysis, we sought to ascertain whether an experimental variation of the uncertainty of objective color parameters (erasure condition = complete uncertainty vs. non-erasure condition = complete certainty) influenced the uncertainty of subjective brightness ratings (high vs. low variations in subjective brightness ratings). A Bayesian independent samples t-test was conducted to examine the effect of experimental condition (erasure vs. non-erasure of objective color parameters) on mean variation of subjective brightness ratings. The results indicated that the evidence was inconclusive, with anecdotal support for the null hypothesis. No statistically relevant difference was observed in the mean variation of subjective brightness ratings between the non-erasure and the erasure conditions. Therefore, contrary to our initial hypothesis, we were unable to reproduce the findings observed in previous studies with this dependent variable. As Study 3 ruled out a limiting factor of Study 1 and 2, it can be considered a conservative test of the objective color parameter erasure effect on this DV. Study 3 additionally increased the number of colors therefore minimizing the explanatory power of the alternative explanation. This fact could have played a role in the failure to replicate the previous results. Nonetheless, the alternative explanation is still comparably probable even with 12 colors when considering an isolated study. Although the evidence for H0 is only anecdotal and a definitive conclusion cannot be drawn from the data, we feel that for this reason and the fact that the mean scores for this DV in both experimental conditions exhibited an opposite trend to that observed previously, serious doubts about the validity of the previous findings seem reasonable. This implies that the evidence found for H1 in the two preceding studies may have been false positives, potentially caused by coincidentally favorable color assignments.

With regard to the result obtained with our second DV, namely the subjective likability of the color seen, a clear confirmation of H1 with extreme Bayesian evidence was found. A Bayesian independent samples t-test was conducted with the experimental condition (erasure vs. non-erasure of objective color parameters) as IV and mean likability rating as DV. The results again demonstrated extreme evidence for H1. The color presented was rated as more likeable in the non-erasure condition than in the erasure condition, which replicated the effects observed with this DV in Studies 1 and 2. This provides confirmatory evidence of the hypothesis that erasure of color parameters affects subjective likability mean scores. Study 3 constituted a more conservative test of this effect than that reported in Studies 1 and 2. This was achieved by eliminating the previously existing confound in random color selection procedures and by using an increased set of color stimuli. In sum, the two exploratory results reported in the two previous studies and the one confirmatory finding reported in Study 3 indicate a preference for colors when objective color parameters are stored, as opposed to a condition in which the objective color parameters are erased. This measurement-dependent variation of subjective color evaluations was observed across three different studies with different sets of randomly selected colors. Although with repeated replications the alternative explanation that the results are biased due to the specific random color selection process becomes more and more unlikely, the alternative explanation is still a valid argument when considered for each study separately. Before overall analyses will be provided to determine the exact nature of the erasure effects found so far and potentially ruling out the biased color argument, we decided to further replicate the effect with a different set of 12 colors. Specifically, the objective of Study 4 was to confirm the experimental manipulation effect on subjective likability observed in Studies 1 to 3 through a final preregistered replication study. The subjective variation of brightness results were deemed not replicable and were not subjected to further testing in a confirmatory manner in the Study 4.

Study 4

Study 4 was preregistered at OSF and designed to replicate the experimental manipulation effect on subjective likability, as documented in a confirmatory fashion in Study 3, and as previously identified in an exploratory manner in Studies 1 and 2. The experimental design was identical to that of Study 3, with the exception that 12 new, different colors were randomly selected as experimental stimuli. The objective of Study 4 was to directly replicate the color parameter erasure effect with the likability mean score as DV found in Study 3. The preregistration can be accessed at OSF (https://osf.io/cvwgq).

Methods

Ethical Guidelines

At the beginning of the online experiment, participants were informed about the study in general terms. The voluntary nature of their participation and the data protection regulations were emphasized. Participants could accept participation by clicking a button (informed consent procedure). All stored data were encrypted and analyzed anonymously. Ethical approval for the study was obtained from an ethics committee.

Sample

The same Bayesian sequential analysis procedures as those employed in Studies 1, 2, and 3, were used to analyze the data, with a minimum evidential criterion of BF ≥ 10 (strong evidence) serving as stopping rule in the main analysis. If this was the case, data collection was terminated and the final BFs were subsequently reported. In the event that a BF ≥ 10 was not attained within a maximum number of participants, i. e., n = 2000, data collection would also have been terminated and the final results would have been reported (Bayesian sequential design with maximum N). Both stopping rules were preregistered.

The participants were recruited primarily through private contacts and social media by LMU students in the context of experimental courses and bachelor's theses (see acknowledgements) 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.

A total of 1,443 participants participated in the study. As defined a priori, participants were excluded from the data analyses if they indicated that their responses were not reliable ("data integrity item," n = 41) or if they indicated that they were color blind (n = 37). The final sample included in the analyses consisted of N = 1,367 participants. The study was deemed complete at this sample size since the BF exceeded 10, as indicated by the Bayesian analysis.

The sample consisted of 940 female, 419 male, and 8 diverse participants. The mean age of the participants was 37.58 years (SD = 12.75; max. age: 84). Data collection was conducted online via PC, tablet, or smartphone, and participants completed the study in German, Spanish, French or English language.

Materials

Colors Used in the Experiment: Objective Color Parameters

In advance of the experiment, 12 distinct colors were randomly selected and assigned to either the erasure or non-erasure condition prior to data collection, serving as experimental stimuli in this study. The random color selection process followed the same protocol as in Study 3. Consequently, both color subsets were selected at random with regard to hue (0 to 359) and lightness (30% to 70%). Once more, saturation was preset at 100% and maintained at a constant level for all 12 colors. The objective color parameters were not examined during this process or while data collection was being conducted.

The objective color parameters of the non-erasure condition were stored in a result file and subsequently inspected. The actual color parameters were as follows: hsl(359, 100%, 60%), hsl(51, 100%, 54%), hsl(11, 100%, 61%), hsl(333, 100%, 56%), hsl(167, 100%, 30%), and hsl(109, 100%, 67%). The objective color parameters of the six colors utilized in the erasure condition were not stored, and the file containing the data was deleted after the conclusion of data collection without undergoing any form of inspection to ensure that the information could never be identified. The hue of the erasure condition colors exhibited random variation between 0 and 359, while the preset saturation was 100% for all colors and the lightness varied between 30% and 70%. This resulted in an uncertainty regarding the objective hue and objective lightness of the experimental color stimuli.

During data collection each participant was randomly assigned to one of the 12 colors. This color was presented as a rectangle (400 by 400 pixels; size dependent on screen used) in the center of the screen for subjective color assessments.

Subjective Color Assessment

The color presented to each participant was subjectively rated on two characteristics: brightness and likability. These ratings were made on visual analogue scales ranging from 0 to 100 as was done in Studies 1, 2, and 3 (for details see corresponding section of Study 1).

Procedure, Design and Statistical analyses

The methodology and design were identical to those employed in Study 3. The statistical main analysis employed a distinct Cauchy prior in comparison to the prior utilized in previous analyses. The effect size estimation was conservative, with a $d_{cohen} = .2$, as observed in a previous study (Study 1). An uninformed prior was selected given the considerable heterogeneity in effect sizes observed across the previous studies. This resulted in a lack of sufficient empirical evidence to inform the prior. In addition, the GQT in its current form did not provide any effect size predictions on theoretical grounds. The uninformed prior Cauchy distribution used was centered around 0 with $r = .2$ (i.e., $\delta \sim$ Cauchy [0, .2]), and a one-sided testing approach was employed (both preregistered).

Results

The main hypothesis was that the mean score of subjective likability of the colors seen would be higher in the experimental condition in which the objective color values were stored (non-erasure condition) than in a condition in which those objective color data were deleted (eraser condition). To test this directed hypothesis, a one-tailed Bayesian independent samples t-test with erasure manipulation (erasure vs. non-erasure) as the IV and mean likability rating as the DV was performed.[2] The independent samples Bayesian t-test (one-tailed, *N* = 1,367) yielded a final BF_{10} = 10.19, indicating strong evidence in support of H1 (d_{cohen} = .14). The BF first reached the threshold of BF ≥ 10 at *n* = 1,364 (see Figure 4a). As anticipated, the mean likability score was higher in the non-erasure group (*M* = 56.16, *SD* = 28.55, *n* = 697) than in the erasure group (*M* = 51.97, *SD* = 29.32, *n* = 670).

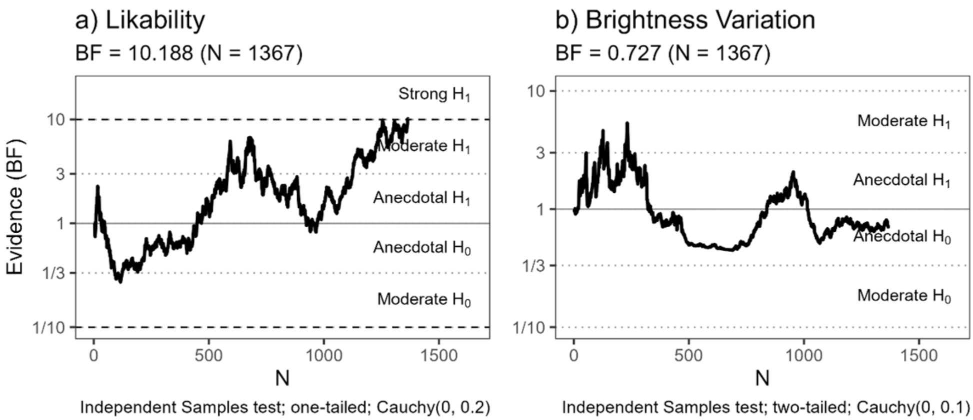
Exploratory Analysis

Additionally, the preregistration included an exploratory analysis of the subjective brightness rating. Specifically, we investigated the impact of erasure on the mean variation of subjective brightness ratings, as previously conducted in Studies 1 to 3. Initially, an erasure effect was identified with this dependent variable (DV) in Studies 1 and 2; however, this result could not be replicated in Study 3. In light of these initial findings, we deemed them to be false positives and thus excluded this DV from our main hypothesis in Study 4. Nevertheless, we conducted an exploratory analysis to test the hypothesis that erasure manipulation of the objective color parameters data (erasure vs. non-erasure) affects the mean variation of subjective brightness ratings (undirected hypothesis). A two-sided independent samples Bayesian t-test was conducted with the experimental condition (erasure vs. non-erasure) as IV and the mean variation of subjective brightness ratings as DV (uninformed Cauchy prior centered at 0 with *r* = .1, i. e., δ ~ Cauchy [0, .1], as in the original analyses). The Bayesian independent sample t-test (two-tailed, *N* = 1,367) yielded a final BF_{01} = 1.38, indicating anecdotal evidence in support of H0 (d_{cohen} = .08; see Figure 4b). The mean variation of brightness score was higher in the non-erasure group (*M* = 14.35, *SD* = 10.84, *n* = 697) than in the erasure group (*M* = 13.57, *SD* = 9.99, *n* = 670). This finding did not replicate the effects observed in Studies 1 and 2, as anticipated.

2 In the preregistration, the hypothesis was clearly formulated and declared as directed, and a "one-sided testing approach" was announced. However, in the subsequent analysis description, a typographical error occurred, resulting in the mention of a "two-tailed" test instead of the intended "one-tailed" test. The exact analyses description section subsequently corrected this error by accurately describing the test as "one-sided."

Figure 4

Accumulation of Bayesian Evidence for the Confirmatory Hypothesis (a) and the Exploratory Hypothesis (b) of Study 4



Discussion

The objective of Study 4 was to directly replicate in a preregistered manner, the color parameter erasure effect on subjective likeness rating mean scores observed in Study 3 (and identified in an exploratory manner in Studies 1 and 2) with a distinct set of 12 color stimuli and once more with a comprehensive random color selection procedure for both conditions (see Study 3). In the main analysis, the Bayesian independent samples t-test (one-tailed) with experimental condition (erasure vs. non-erasure of objective color parameters) as IV and mean likability rating as DV revealed strong evidence for H1. As anticipated, the mean likability rating for the color presented in the non-erasure condition was higher than that for the erasure condition. This directly replicated the effect observed with this DV in Study 3. This provided further confirmation of the existence of an experimental manipulation effect on subjective likability mean scores. Study 4 constituted the most rigorous test of the hypothesis, as it was preregistered and focused exclusively on this DV. This effect was observed with a new set of 12 color stimuli. In conclusion, the two exploratory results reported in the two initial studies and the one confirmatory finding reported in Study 3 were again documented in Study 4, indicating a replicable experimental manipulation effect on subjective likability across four different studies. Also, in this study a clear interpretation of the effect found cannot be provided since the effect of the erasure-manipulation can also be attributed to a biased random color subsets selection that favored the confirmation of H1 by chance.

Our research agenda reported here involved three separate steps. First, subjective experiences need to vary with the erasure manipulation consistently in each study (Step 1). This goal was achieved and an experimental manipulation effect on subjective likability was documented in each single study. Second, such findings need to be shown to be replicable across several studies (Step 2). This was also the case as reported above. And third, the alternative factor (the biased color argument) needs to be ruled out by final overall analyses (Step 3) to rectify the internal validity violation within the experimental designs allowing to clearly attribute the experimental effects found so far to the erasure manipulation. This three-step research agenda will next be completed by reporting the results of the overall analyses. The focus will be on the likability of color rating, since only for this DV Steps 1 and 2 were successfully completed.

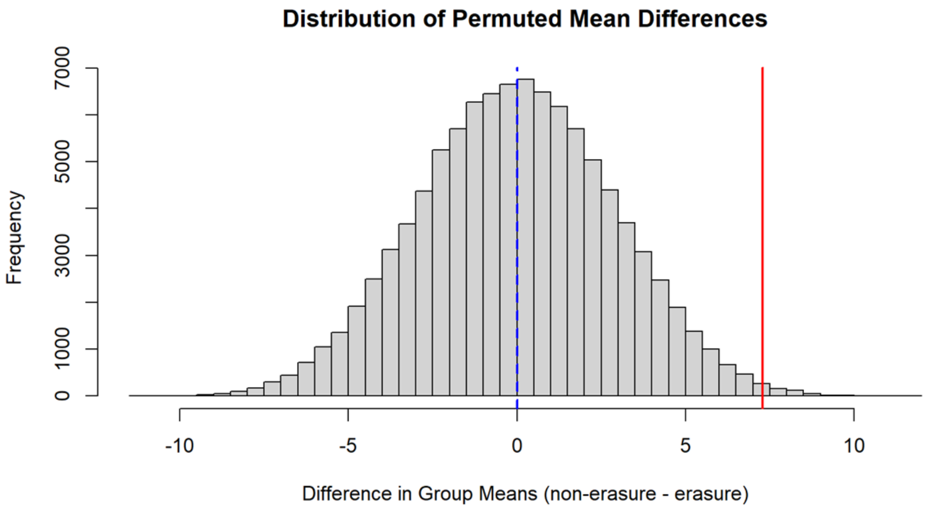
Overall Analyses for Likability: Testing the Impact of Randomly Biased Color Selections and the "Lightness Range" Confound

The status of our findings across four studies is the following: An experimental manipulation effect was found consistently for one DV, the subjective likability of color. This effect appeared robust across four independent studies. However, the erasure effect interpretation was intentionally confounded in each study by a random assignment of limited subsets of colors (3 and 3 in Studies 1 and 2; 6 and 6 in Studies 3 and 4). This weakened the internal validity of each study and made it impossible to attribute the effects found solely to the erasure manipulation. Rather, for each study biased ("lucky") color assignments to experimental conditions can alternatively explain the effects. Following our research agenda, in Step 3 now this alternative explanation will be submitted to frequentist statistical tests. In addition, the lightness range confound present in Studies 1 and 2 (see discussion section of Study 2) will also be addressed in one overall analysis. In preparation of this, all data of the four studies were combined into one complete data file (n_{total} = 3,623).

The primary overall analysis, a permutation analysis, was a priori planned before data collection. It was performed after data collections of all studies were completed to test the potential impact of biased color selections on the erasure effects found in the studies. Across all studies, the erasure-manipulation increased the likability ratings on average by 7.28 points (mean likability score non-erasure [$M_{non\text{-}erasure}$ = 57.56, $SD_{non\text{-}erasure}$ = 27.13] minus mean likability score erasure condition [$M_{erasure}$ = 50.28, $SD_{erasure}$ = 28.40]). The permutation analysis directly tested the likelihood to find a mean score difference of 7.28 or greater by chance given the specific 36 colors and their observed likability ratings. The subjective likability weighted-mean scores of all colors used in the studies were randomly permutated and reassigned to the experimental conditions 100,000 times. This data permutation approach revealed that the probability to find a likability mean score difference (likability $M_{\text{non-erasure}}$ minus likability M_{erasure}) of equal or greater than 7.28 was p = .005 (see Figure 5).

Figure 5

Distribution of Condition Mean Differences for 100,000 Permutations and Actual Mean Difference (Red Line)



Since Studies 1 and 2 contained a confound caused by the unequal generation of the colors' lightness scores in each experimental condition, which potentially violated the core assumption of permutation testing that observations are exchangeable under the null hypothesis, the permutation analysis was additionally performed exclusively for Studies 3 and 4, in which this color selection confound was absent. The total difference between experimental groups for both studies was $d_{\text{non-erasure vs. erasure}}$ = 5.74. The subjective likability weighted-mean scores of all colors used in the studies were again randomly permutated and reassigned to the experimental conditions 100,000 times. The permutations revealed that the probability to obtain this or a higher d-score was p = .05.

In sum, the permutation approach provided the most direct test of our specific concern: given these particular colors and their ratings, what is the likelihood that random color assignment alone could produce the observed pattern? The low probability ($p \leq .05$) found in both permutation analyses provided convincing evidence against the "lucky assignment" explanation.[3]

As a supplementary post-hoc analysis, we conducted a multilevel model analysis (MLM) which included the mean likability rating as DV, the erasure manipulation as IV (fixed effect) and the 36 colors used in all studies as random factor. We tested the hypothesis that there was an overall erasure effect in likability ratings when controlling for color-specific variations (H1: likability $M_{\text{non-erasure}} \neq$ likability M_{erasure}; two-sided). The analysis yielded a significant fixed effect of condition, indicating that likability ratings were higher in the non-erasure condition compared to the erasure condition, representing a small-to-medium effect size, $\beta = 7.46$, $t(32) = 2.94$, $p = .006$, $d = .28$; ICC = .06. The low ICC suggests that while there are some differences between study colors, most of the variance in ratings comes from individual responses rather than systematic color effects. A sensitivity analysis confirmed that the effect of condition (red line at 7.28, see Figure 6) remained stable. Excluding individual colors led to only minor fluctuations around the original mean difference, and no single color exclusion dramatically altered the effect. These findings suggest that the erasure manipulation effect was robust and not driven by specific colors.

A second model that included "Study" as an additional random factor revealed that the variance attributed to "Study" was essentially zero (0.00), indicating no meaningful variation between studies. This suggests that the erasure effect was not dependent on the study part and therefore that the differences of color value generation (lightness confound) between Studies 1 and 2, vs. 3 and 4 likely did not contribute to the observed erasure effect on likability ratings and can therefore be considered irrelevant for our interpretation.

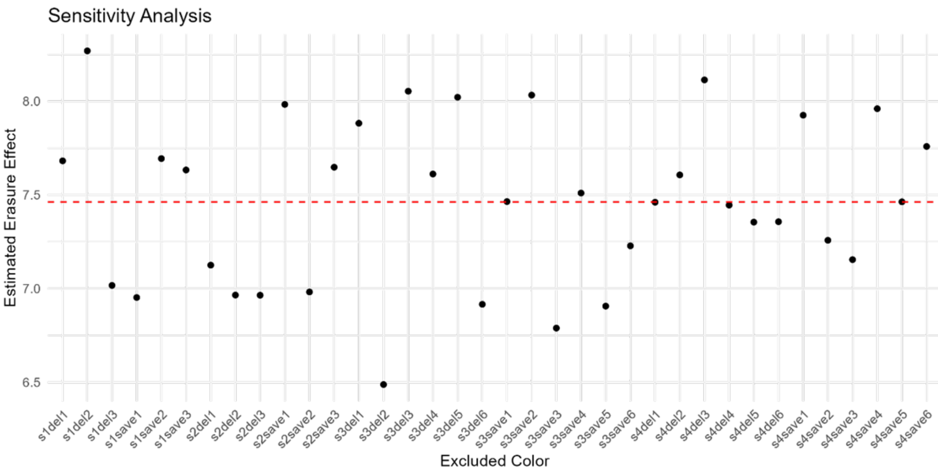
This analysis implicitly assumes that our 36 colors constitute a representative sample of the broader color space. The model can therefore only control for color variance within our specific random sample. If this sample happened to contain colors that were unusually similar in their appeal (due to sampling variability), the MLM might underestimate how much color differences could explain the effect in the broader color space. However, since the permutation analysis showed that even within this specific sample of 36 colors, random assignment is highly unlikely to explain our observed effects ($p \leq .05$), concerns about sample representativeness do not under-

3 It can be argued that while there is no issue in expanding Bayesian evidence for individual studies 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 up until the first time BF ≥ 10 was met and submitted them to a further permutation analysis. The resulting sample consisted of $n = 2{,}453$ participants and showed a $d = 5.94$ between conditions yielding $p = .03$. Considering only Studies 3 and 4, the subsample consisted of $n = 1{,}653$, showing a $d = 5.07$, $p = .10$ (see analysis script at https://osf.io/u2zqp).

mine our primary conclusion. All data and overall analyses can be retrieved from the OSF project (https://osf.io/u2zqp/).

Figure 6

This Analysis Shows the Estimated Erasure Effect When Individual Colors Are Dropped



Discussion

In the overall analyses provided here, Step 3 of our stepwise research agenda was performed. These analyses addressed the problem of a non-ambiguous causal interpretation of the experimental erasure manipulation on likability ratings by ruling out the main alternative explanation of "lucky" random color assignments to experimental conditions (permutation analysis and supplementary MLM) and the "lightness range" confound present in Studies 1 and 2 (additional MLM with study as random factor).

The a priori planned 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 our specific 36 colors was very low (p = .005 for all studies; p = .05 for Studies 3 and 4 alone). By testing whether random re-assignments of these exact colors and their observed ratings could produce the observed pattern, this analysis directly addressed the "lucky assignment" alternative explanation.

The supplementary MLM analysis (post-hoc) further supported the robustness of the experimental erasure effect. The results showed that the significant effect of the erasure manipulation (β = 7.46 points) existed after controlling for color-specific variation within our sample. The low ICC (.06) indicated that most variance came from individual responses rather than systematic color effects, and sensitivity analyses confirmed the effect remained stable when individual colors were excluded. However, this analysis assumes that our 36 randomly selected colors adequately represent the broader color space for statistical control purposes. While this assumption cannot be definitively verified, the permutation analysis renders concerns about our specific sample highly unlikely to explain the observed effects. The additional MLM including "study" as a random factor revealed essentially zero study-level variance, confirming that the lightness range differences between Studies 1–2 and Studies 3–4 did not contribute to the erasure effect.

In sum, the permutation analysis provided compelling evidence that our effects were not due to chance ("lucky") color assignments, while the MLM analyses demonstrated robustness within our observed color set and ruled out study-level confounds. Together, these findings support the assumption that the erasure manipulation of objective color parameters itself had an effect on subjective likability ratings. These analyses rectified the internal validity violation within our experimental settings, making an overall causal explanation 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 color assignment to experimental conditions" interpretation. This confound was supposed 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 data set. This procedure followed similar recommendations made by Römer (2024) and Lucadou (2001).

As a consequence, these findings lend support to the conjecture derived from the GQT that macroscopic complementary relations exist between objective and subjective color measurements. The results indicate that colors are perceived as more favorable when the objective color parameters are retained compared to a condition in which they are erased. Step 3 analyses thus provided preliminary 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; Lucadou et al., 2007; Römer, 2023, 2024; Walach & von Stillfried, 2011) which proposed in our experiments that experimental variations in objective color parameters' documentations bias subjective experience of the colors under investigation.

General Discussion

The objective of the studies presented here was to identify evidence for the existence of macroscopic complementary relations between subjective experiences and objective measurements of color features. The non-commutability conjecture of the GQT (Atmanspacher et al., 2002; Walach & Römer, 2000, 2011; see also Fach, 2011; Filk & Römer, 2011; Hinterberger & von Stillfried, 2013; Lucadou et al., 2007; Römer, 2023, 2024; Walach & von Stillfried, 2011) posits that this kind of psychophysical interaction entails a change in the state of the whole system, including the eigenvalues of the subjective color experience subsystem, as a result of an act of measurement of the objective color parameter subsystem. Specifically, it was expected that an experimental erasure manipulation of the objective color parameters hue and lightness would affect changes in variations of subjective brightness ratings and in subjective likability mean score ratings of the colors presented.

The acausal nature of this relationship under investigation required a deviation from the commonly accepted scientific standards of testing such effects. As mentioned at the end of the introduction section and throughout the article, when acausal macroscopic complementary relations are subjected to causal, experimental tests, unambiguous causality test might ultimately destroy the acausal nature of the effect under study making it impossible to find robust evidence for it (for similar arguments, see Lucadou, 2001; Römer, 2024). Thus, a second alternative explanation was introduced into the design, the "biased color assignment to experimental conditions" argument, which was implemented by a limitation of the number of colors used in the studies. In this way, on the study-level a clear causal interpretation of an observed experimental manipulation effect was obscured, leading to two concurring interpretations. Firstly, the observed changes in subjective likability depending on whether a measurement of the objective color parameters was carried out or not could be conceptualized as a macroscopic complementary relation between subjective and objective aspects of a phenomenon. It would appear that objectively measured realities are more readily subjectively assessed as likeable than non-objective ones. One possible explanation is that psychological processes, such as familiarity with objective worldviews, facilitate a preference for objectively documented phenomena. Secondly, since each study operated with limited color sets, an incidentally favorable color assignment cannot be ruled out at this stage. Importantly, it was assumed that for a definitive and reliable effect documentation to be achieved (Step 1 and 2), it was necessary that both causal mechanisms could simultaneously account for the results within a study. Consequently, the coexistence of both explanations was a crucial element of the experimental setting and served as an indispensable prerequisite for Step 1 and 2. After all studies have been performed, Step 3 then disentangled the confound and retrospectively provided evidence for the underlying causal mechanism affecting subjective likability ratings.

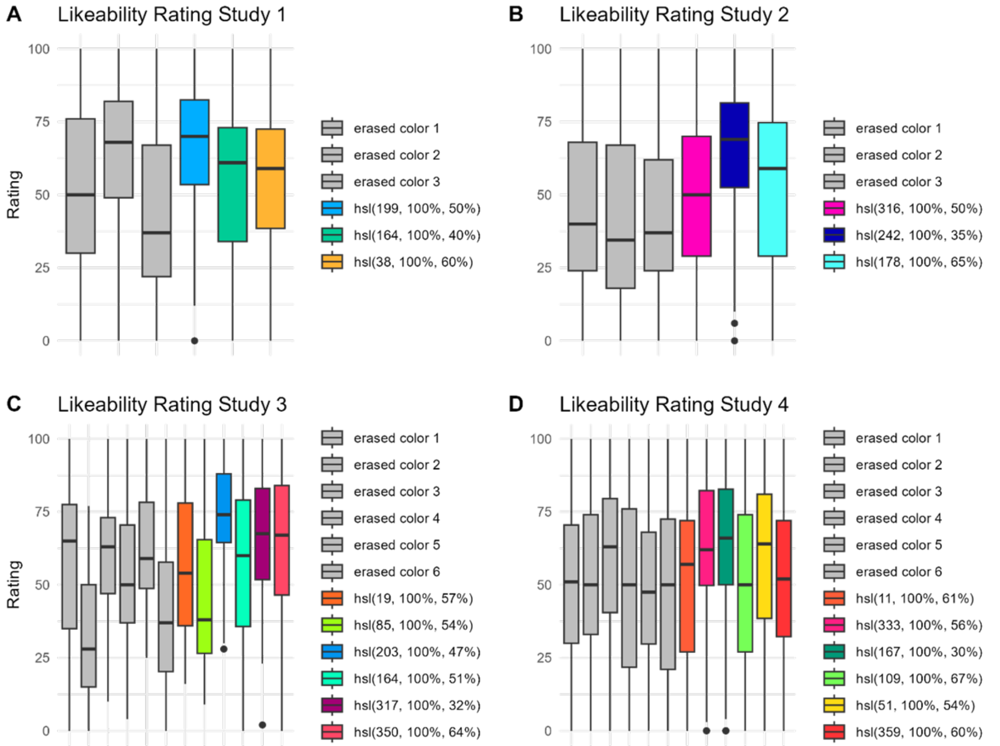
With regard to Steps 1 and 2, experimental manipulation effects on variations of subjective brightness ratings were only found in two studies, but the effects were not robust across all four studies. Study 1 yielded strong evidence supporting H1, indicating higher variations in subjective brightness scores when objective color parameters were not erased compared to the erasure condition. This finding is analogous to the Heisenberg uncertainty relation (Heisenberg, 1927). Although this finding was replicated in Study 2, a further confirmatory test of the hypothesis was unsuccessful in Study 3 and an exploratory test of the hypothesis also failed in Study 4. In conclusion, the initial promising results should be regarded as false positives, as no replicable experimental manipulation effects could be documented with this dependent variable across all studies and the initial effects can be attributed to chance occurrences. Consequently, the Step 3 analyses were not performed.

With regard to Steps 1 and 2, the experimental manipulation effect on subjective likability ratings were detected in an exploratory fashion in both Studies 1 and 2. Subjective likability ratings were higher in the non-erasure condition than in the erasure condition in both studies. These findings were corroborated in a replication attempt in Study 3 and further confirmed in a preregistered replication Study 4 (see Figure 7). In sum, the criteria set and tested in Step 1 and 2 were successfully met with this dependent variable. Next, in Step 3 the overall analyses performed with this dependent variable revealed retrospectively that the experimental manipulation effect on subjective likability ratings could hardly be attributed to biased color assignments to conditions or other confounding factors leaving the erasure manipulation as the most probable factor causing the mean score differences in likability ratings. Thus, after the Step 3 analyses, it can be stated with some confidence, that the presence or absence of objective color parameter documentations had a significant impact on the subjective evaluations of colors. This replicable erasure effect together with the Step 3 analyses provided supportive evidence for complementary relations of subjective color likability experiences and objective assessments of these colors, as predicted by the GQT (Atmanspacher et al., 2002; Walach & Römer, 2000, 2011; see also Fach, 2011; Filk & Römer, 2011; Hinterberger & von Stillfried, 2013; Lucadou et al., 2007; Römer, 2023, 2024; Walach & von Stillfried, 2011).

The question arises as to why a macroscopic complementarity relationship could only be found with the likability of a color and not with the subjectively assessed variations of brightness values? We suspect that the latter is not inherently subjective enough, but rather represents a subjectively clouded approximation of the objective color parameters. Likability, on the other hand, has no corresponding counterpart in the objective parameter space and therefore comes closer to the ideal of a Qualia. Looking back on this series of experiments, we assume that only such subjective perceptual phenomena are suitable for demonstrating a corresponding interaction between the dual aspects of reality.

Figure 7

Color Likability Scores Across All Four Studies



Another critical question might be what the term "color objectification" involved in our design. Participants saw the respective color regardless whether they were assigned to the erasure or to the non-erasure condition. Thus, every color was registered at least by several individuals. Why does this not constitute an objective color assessment? In our view, the participants' perceptions of a color is not an objective documentation of the respective color seen, but would rather constitute an eye-witness testimony if asked about the color. Participants were also unable to communicate their perceptions across each other, since only individuals were anonymously tested. Given the solipsistic nature of their color perceptions, they do not qualify as an objective color documentation. Rather on the experimenter level, the color parameters stored for the non-erased colors were inspected by experimenters later on and can in the future be verified by anyone who likes. This should in line with the GQT (Römer, 2024) be considered the highest standard of objective color parameters report that can be reached in a scientific setting and is thus the true origin of the erasure-effects reported here.

The non-local nature of these interactions is supported by the fact that participants were unaware of the erasure manipulation and that objective data were stored or deleted after subjective responses were assessed. Therefore, the results cannot be explained by classical-local information transfer and should be considered non-local. These complementary relations were found to be robust and replicable across four studies, which lends support to our initial assumption that structural correlations are robust and replicable phenomena when tested in an acausal manner on the study level (Lucadou, 2001; Römer, 2024). This is in line with the proposal by Pauli and Jung in their mind-matter framework (see Atmanspacher & Rickles, 2022). The data indicate that psychophysical interactions between subjective and objective elements involved in color assessments are based on acausal and non-local entanglement correlations that are not limited by the NT axiom when strict causality tests are avoided (Lucadou et al., 2007; Römer, 2024). This is in contrast to induced correlations. They require non-local entanglement correlations to be transferred into classical causal interrelations between the subjective and objective realms, which are evident in normative volitional processes (Maier et al., 2022). This breaking up of a non-locally entangled system into a classical causal system might result in an effect and decline data pattern, as specified by the NT axiom. Causal mechanisms when induced correlations are studied are already present at the participant level in addition to the study level when testing robustness and replicability. They thus inherently stand in conflict with the NT axiom. With regard to the structural correlations tested here, the acausal, non-local entanglement correlations remained intact upon scientific documentation since no causal mechanisms were initiated neither at the level of participants' actions (since no normative volitional force was present) nor – since strict causality tests were avoided – at the study-level. Thus, the effect-decline pattern (Lucadou et al., 2007; Römer, 2024) could be circumvented. It remains to be seen whether future successful replications of Studies 3 and 4 can be expected based on the existing data and after Step 3 analyses have been performed. In relation to this, one critical aspect should be highlighted here. We argued that due to the acausal nature of the phenomenon under study, a strict causality test obtained from internally valid experiments was considered to ultimately fail. However, someone might object that a final causal test was provided in the overall analyses and this might contradict the proposed impossibility to causally test the erasure effects. We disagree with this argument. As mentioned above, strict causal tests at the study level when testing acausal psychophysical interactions are prohibited by the conjectures of the GQT (Römer, 2024). However, this restriction might not include weaker causality tests such as those provided here. The overall analyses just re-established a causal interpretation of the data in a retrospective manner. This deviated from standard scientific approaches when testing causality and might be considered an unusual but feasible alternative which circumvents the scientific causality testing paradox inherent in GQT (see also Grote, 2017; Lucadou, 2001; Lucadou et al., 2007; Walach et al., 2022). This strategy was supposed to assure the occurrence of an erasure effect within several consecutive studies without violating the NT axiom of the GQT (Römer, 2024). It remains open whether such an indirect causality test could be replicated in the future. The actual findings are thus only preliminary.

Limitations

In the following, several limitations of the research presented here need to be highlighted and the interpretation of the overall results should be considered in light of them.

First, it should be noted that the use of a parametric testing approach may be subject to certain limitations with regard to the subsequent data analysis. Parametric testing with independent sample t-tests is predicated on two fundamental assumptions: homogeneity of variances and normality of the data tested. With regard to the likability ratings, the homogeneity assumption was fulfilled. In accordance with the recommendations of Lumley et al. (2002), violations of normality were disregarded due to the considerable sample sizes employed in these studies. Their findings indicated that independent sample t-tests remain robust against such violations when sample sizes exceed 500. Therefore, our a priori selected parametric testing approach was deemed appropriate. However, van Doorn et al. (2021) argue that violations of normality are detrimental to the reliability of independent sample t-tests regardless of sample size. They suggest that non-parametric testing may be a more appropriate alternative in such cases. Although this issue remains a topic of contention (Knief & Forstmeier, 2021) and the community appears to be divided in opinion about this matter, we decided to additionally assess the normality of our data and incorporated non-parametric tests in instances where this assumption was breached to inform those who prefer a more conservative approach. Q-Q plots of the likability data indicated that the normality assumption was violated in all studies. Consequently, we conducted supplementary analyses to assess the impact of the erasure manipulation effect on mean likability scores using non-parametric Bayesian and frequentist Mann-Whitney-U tests. The BF_{10} and p-values obtained from these analyses were: Study 1, $BF_{10} = 6.79$; $p = .003$ (two-tailed); Study 2, $BF_{10} = 38679.32$; $p < .001$ (two-tailed); Study 3, $BF_{10} = 256.89$; $p < .001$ (two-tailed) and Study 4, $BF_{10} = 6.71$; $p = .003$ (one-tailed). These findings, though somewhat weaker than those obtained from parametric testing, largely corroborated the results obtained from the parametric independent sample t-tests.

Second, in Studies 1 and 2, the experimental condition was confounded with different lightness ranges from which the colors were randomly selected. This could alternatively explain the erasure effects on the DVs. With regard to likability this seems rather unlikely since in Study 3 and 4 this confound was eliminated and still erasure effects were documented with this DV. In addition, a MLM analysis combining the data sets of all four studies tested a random effect of study. The results indicated that variations across studies did not affect the erasure effects found on likability ratings. However, this leaves open the question whether and how subjective brightness ratings were affected by this confound.

Finally, all studies were conducted online. This allowed to collect data with high-power samples sizes. A potential downside of this is that the color stimuli were presented under different viewing conditions. Viewing conditions might play a crucial role in subjective brightness and likability assessments. Thus, online studies probably increased the error variance leading to an underestimation of the real effect sizes. Although, we think that sample power compensated for this random source of individual variance, in future studies more controlled settings including lab studies should be considered as alternatives.

Conclusion

In conclusion, despite the limitations reported above the erasure effect on likability found in the overall analyses (Step 3) of the four studies provided a preliminary successful test of the conjectures derived from the GQT (Atmanspacher et al., 2002; Walach & Römer, 2000, 2011; see also Fach, 2011; Filk & Römer, 2011; Hinterberger & von Stillfried, 2013; Lucadou et al., 2007; Römer, 2023, 2024; Walach & von Stillfried, 2011). This theory postulates that there are macroscopic complementary relations between subjective and objective elements within a psychophysical system, such as color experience. The relation can be described by means of non-local entanglement correlations in case the measurements within the subsystems are spatio-temporally separated, which was the case in this research described here, and can be subjected to systematic study. Furthermore, the theory posits that psychophysical realities are observation-dependent, leading to the non-commutability conjecture. This states that measurements performed on one subsystem of a psychophysical system result in changes to observables derived from the other complementary related subsystem. All these conjectures were confirmed in our studies and their overall analyses for one of the DVs. This constitutes initial evidence for the existence of macroscopic complementary relations as proposed by the GQT.

The ontic separation and causal connectedness between subjective and objective aspects of macroscopic reality, as proposed by Descartes' substance-dualism (Descartes, 1641), and the complete reduction of subjective phenomena to physical concepts, as put forth in physicalism (Stoljar, 2024), both appear to be overly simplistic when considering our results. Rather, our findings indicate that the subjective and objective are dual aspects of reality that exist in an entangled unity. In other words, the two aspects of reality are complementary and interdependent. Consequently, the data are incompatible with physicalism, which posits that the subjective and objective are both purely physical in nature. In purely physical systems, however, complementary relations between macroscopic subsystems – in our case, ratings of subjective likability and assessments of objective color parameters – are considered impossible due to Planck's constant (DeBroglie, 1958; Römer, 2024) and decoherence (Zeh, 2012). It is only if the subjective is conceived as a non-physical aspect of the world that macroscopic complementarity phenomena, as documented in the overall analyses of the four studies presented here, can occur. Furthermore, our findings also address a fundamental tenet of Descartes' substance dualism, namely that causal processes exclusively underly the interrelationship between the subjective and the objective substance. Rather, dualistic approaches should consider acausal, non-local entanglement correlations as additional or even alternative explanations for these interrelations, given the arguments raised against the compatibility of substance dualism with inter-substance causality (Chalmers, 1995; Levine, 1983; Nagel, 1974, Shariff et al., 2008). Finally, our findings align with the tenets of dual-aspect monism (see Atmanspacher & Rickles, 2022), which postulate a unified reality wherein 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.

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

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 physics*. Cambridge University Press.

Bohr, N., & Favrholdt, D. (1999). *Collected works. Volume 10: Complementarity beyond physics (1928–1962)* (ed. David Favrholdt). Elsevier.

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

DeBroglie, L. (1958). *Licht und Materie*. Fischer.

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, viâ Iacobeâ, sub signo Phœnicis.

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

Fechner, G. T. (1860). *Elemente der Psychophysik* (Erster Teil). Breitkopf & Härtel.

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

Furnham, A. (2001). Self-estimates of intelligence: Culture and gender difference in self and other estimates of both general (g) and multiple intelligences. *Personality and Individual Differences, 31*(8), 1381–1405. https://doi.org/10.1016/S0191-8869(00)00232-4

Gignac, G. E., & Zajenkowski, M. (2019). People tend to overestimate their romantic partner's intelligence even more than their own. *Intelligence, 73*, 41–51. https://doi.org/10.1016/j.intell.2019.01.004

Grote, H. (2017). Multiple-analysis correlation study between human psychological variables and binary random events. *Journal of Scientific Exploration, 31*, 231–254.

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

Knief, U., & Forstmeier, W. (2021). Violating the normality assumption may be the lesser of two evils. *Behavioral Research Methods, 53*(6), 2576–2590. https://doi.org/10.3758/s13428-021-01587-5

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly, 64*, 354–361. https://doi.org/10.1111/j.1468-0114.1983.tb00207.x

Lucadou, W. v. (2001). Hans in luck: The currency of evidence in parapsychology. *Journal of Parapsychology, 65,* 3–16.

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.

Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual review of public health*, *23*(1), 151–169. https://doi.org/https://doi.org/10.1146/annurev.publhealth.23.100901.140546

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., Dechamps, M. C., & Rabeyron, T. (submitted). *Effect of quantum-double-slit outcomes' erasure on subjective recall of those outcomes – A failed replication*. Manuscript submitted for publication.

Maier, M. A., Geissler, C., DeMattia, E. C., Vogel, A., & Dechamps, M. C. (2024). Complementarity of subjective and objective realities: An experimental investigation of variations of subjective self-reported intelligence data when objective data are erased or not-erased. *Journal of Anomalistics, 24,* 15–54. https://doi.org/10.23793/zfa.2024.015

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

Neubauer, A. C., & Hofer, G. (2020). Self- and other-estimates of intelligence. In R. J. Sternberg (Ed.), *Cambridge Handbook of Intelligence* (2nd ed., pp. 1179–1200). Cambridge University Press. https://doi.org/10.1017/9781108770422.050

Palmer, S. E. & Schloss, K. B. (2010). An ecological valence theory of human color preference, *Proceedings of the National Acadademy of Sciences, 107*, 8877–8882. https://doi.org/10.1073/pnas.0906172107

Perler, D. (2006). *René Descartes*. Beck.

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.

Ruffing, E. (2021). *Einführung in die Geschichte der Philosophie* [Introduction into the History of Philosophy]. (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* (Spring 2024 Edition). https://plato.stanford.edu/archives/spr2024/entries/physicalism/

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., & Haaf, J. M. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*, *28*, 813–826. https://doi.org/https://doi.org/10.3758/s13423-020-01798-5

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

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

Zeh, H.-D. (1970). On the interpretation of measurement in quantum theory. *Foundations of Physics*, *1*, 69–76. https://doi.org/10.1007/BF00708656

Zeh, H.-D. (2012). *Physik ohne Realität: Tiefsinn oder Wahnsinn?* Springer. https://doi.org/10.1007/978-3-642-21890-3

Deutsches Abstract

Makroskopische komplementäre Beziehung zwischen subjektiven Beobachtungen und objektiven Messungen von Farben

In der Verallgemeinerten Quantentheorie (GQT), einer Theorie zur Beschreibung psychophysischer Phänomene, werden subjektive und objektive Aspekte der Realität als komplementär betrachtet. Die hier vorgestellte Forschung hatte das Ziel, eine solche makroskopische Komplementarität zwischen subjektiven und objektiven Realbeschreibungen im Kontext der Farbbeurteilung zu untersuchen. Insbesondere stellt die Nicht-Kommutativitätsannahme der GQT die These auf, dass zwei Observablen, die aus den subjektiven und objektiven Teilsystemen stammen, nicht gleichzeitig spezifische Eigenwerte liefern. Vielmehr verändert der Messakt innerhalb eines Teilsystems, z. B. die Durchführung objektiver Messungen an einem Stimulus, den Zustand des gesamten Systems, einschließlich der Eigenwerte des anderen Teilsystems, z. B. des subjektiven Erlebens dieses Stimulus. Diese Vermutung wurde empirisch in vier Studien und in drei Gesamtanalysen des vollständigen Daten-

satzes über alle vier Studien getestet. Die experimentelle Manipulation beinhaltete eine Messvariation eines Aspekts des vermeintlichen komplementären Paares, nämlich die Speicherung (Nicht-Löschungsbedingung) oder Löschung (Löschungsbedingung) objektiver Farbparameter (Farbton und Helligkeit). Es wurde erwartet, dass die Löschungsmanipulation den anderen Teil des Paares beeinflussen würde, nämlich die subjektiven Bewertungen der Farbe in Bezug auf Helligkeit und Gefallen. Die primäre Hypothese, die löschungsabhängige Effekte auf Variationen der subjektiven Helligkeitswerte testete, wurde in den Studien 1 und 2 bestätigt, konnte jedoch in den Studien 3 und 4 nicht repliziert werden. Die anfänglichen Ergebnisse, die mit dieser abhängigen Variablen gewonnen wurden, können daher als False-Positives gewertet werden, und die primäre Hypothese konnte nicht bestätigt werden. Die zweite Hypothese, die Löschungseffekte auf die mittleren Werte des subjektiven Gefallens testete, ergab folgende Ergebnisse: Die explorativen Erkenntnisse aus den Studien 1 und 2 sowie die bestätigenden Ergebnisse aus Studie 3 und der präregistrierten Studie 4 zeigten einen Löschungseffekt auf das subjektive Gefallen, wobei in jeder Studie höhere Gefallenswerte in der Nicht-Löschungsbedingung im Vergleich zur Löschungsbedingung beobachtet wurden. Da das experimentelle Design absichtlich durch eine alternative „Farbverzerrung“ kontaminiert wurde, um die Effektdokumentation zu stabilisieren, können keine eindeutigen kausalen Löschungseffekte auf Einzelstudienebene nachgewiesen werden. Um diese Kontamination zu überwinden, wurde eine Gesamtanalyse-Strategie vorab geplant und an dem kombinierten Datensatz durchgeführt. Mehrere Gesamtanalysen schlossen die alternative Erklärung der „Farbverzerrung“ und eine weitere Störvariable aus. Zusammenfassend lieferten die Gesamtergebnisse Belege für einen Löschungseffekt auf das subjektive Gefallen, die das Modell der GQT stützen. Dies kann als makroskopische komplementäre Beziehungen zwischen objektiven und subjektiven Realbeschreibungen interpretiert werden, wie sie von der GQT vorgeschlagen werden. Die Implikationen dieser Ergebnisse für unser Verständnis der Natur der Realität und für die Gültigkeit der GQT werden diskutiert.