Unreal? A Behavioral, Physiological, and Computational Model of the Sense of Reality

Gadi Drori [a1], Paz Bar-Tal [a1], Oded Hirsh [a], Ariel Berlinger [b], Niva Goldberg [a], Yair Zvilichovsky [a], Uri Hertz [c], Adam Zaidel [a], Roy Salomon [c, b ,d]

[a] Gonda Brain Research Center, Bar-Ilan University, Ramat Gan, Israel

[b] School of Psychological Sciences, University of Haifa, Haifa, Israel

[c] Department of Cognitive Sciences University of Haifa, Haifa, Israel

[d] The Integrated Brain and Behavior Research Center (IBBRC), University of Haifa, Haifa, Israel

Abstract

An intriguing aspect of the human mind is our knowledge that our perceptions may be false. Our frequent exposure to non-veridical perceptions such as those found in dreams, illusions and hallucinations cause us to examine the actuality of our sensory experiences. As such humans continuously monitor the veridicality of their perceptions in a process termed the Sense of Reality (SoR). Moreover, the Sense of Reality is a central criterion in the assessment of neurological and psychiatric health. The scientific study of hallucinatory experiences has been hindered by their transitory, subjective, ineffable nature and by the fact that they typically co-occur with psychiatric, medical or psychedelic states. Despite the critical role of SoR in daily life and in pathologies little is known regarding its cognitive, physiological and computational underpinnings. Here we employed a novel immersive virtual reality paradigm to induce *Virtual Hallucinations* (VH) simulating the phenomenology of visual hallucinations found in psychiatric, neurological, and pharmacological conditions. Combining psychophysics, physiological recordings and computational modeling in one exploratory (n = 31) and one preregistered experiment (n = 32) we examined responses to VH of varying magnitudes and domains. Judgments of SoR varied depending on the magnitude and domain of VH. These were accompanied by distinct motor, pupillometric and cardiac responses. Finally, sense of reality can be well explained by a computational model in which decisions of reality are based on comparison of current sensory experience to an internal model of the world. Our results shed some light on the age-old question 'how do we know what is real'?

Introduction

Our ability to function and survive in the world depends on a close correspondence between our representation of the environment and its actual characteristics e.g.,[1,2 but see 3]. This correspondence is often termed 'reality', and can be contrasted with other states such as dreams and hallucinations which often feel real despite being non-veridical. Thus, while we typically trust our sensory experiences, we are also aware that they may not be a veracious representation of the world[4]. As such humans continuously monitor the veridicality of their perceptions in a process termed the Sense of Reality (SoR).

The Sense of Reality (SoR) is a central criterion in the assessment of neurological and psychiatric health[5–7]. Distortions of perceptual reality in the form of hallucinations originating from neurological, psychiatric and pharmacological origins are commonplace[8]. For example, around 40% of Parkinson's disease, 27% of schizophrenia patients and between 10-30% of the general population may experience visual hallucinations[9].

Intriguingly, hallucinations can be experienced with or without insight into their nature as false perceptions. For example, patients with visual hallucinations due to macular degeneration (i.e., Charles Bonnet syndrome) typically identify these perceptions as hallucinations[19]. Contrarily, psychotic patients often lose the ability to discern between real and hallucinatory percepts, which has been linked to poorer prognosis[20] and is associated with reduced cognitive abilities[21,22]. Importantly, lapses in SoR in psychosis have been suggested to arise from a convergence of abnormal perceptual processing and reduced metacognitive monitoring of the reliability of these signals[23,24]. Abnormal SoR may also be relevant to other psychiatric symptoms such as derealization and depersonalization, in which the external world or the self are experienced as strange or unreal[5,25,26] , typically in the absence of hallucinatory experiences. Despite the

obvious clinical significance of SoR, there is insufficient knowledge of the mechanisms supporting its function.

In recent years there is growing evidence suggesting the brain infers the probable causes of incoming sensory information based on learned statistical regularities of the environment[27–29]. Such Bayesian frameworks posit that the brain generates our belief about the source of information - the posterior - based on how well it describes the incoming sensory information (likelihood) - and our prior expectations (prior)[29–31]. When a highly atypical sensory event arises (e.g., a rippling building; See Fig. 1A) we must evaluate if the cause of this experience is "real" (i.e., a rippling building) or is "unreal" (e.g., I'm hallucinating). Importantly, we constantly encounter slightly atypical perceptual information due to variance in environmental (e.g., strange reflections or a road rippling due to heat) and internal (e.g., sensory or neural noise) states. Thus, to infer the source of sensory information we use our prior knowledge of the world (e.g., buildings do not ripple) to decide the probability that this is a veridical perception. Hence, SoR may be considered a process of active inference based on sensory signals and our internal model of the world.

Such causal inference approaches have been successful in modelling many aspects of human perception and behavior[29,30,32,33]. Under this framework, SoR can be viewed as a probabilistic inference expressing the fit between a given sensory signal and our model of the world. For example, sudden changes in the color spectra of the world could be attributed to different causes (e.g., lighting changes, colored glasses) which we commonly encounter (i.e., wide *likelihood* distribution). Thus, changing colors of the visual scene, for which we have more experience, should have lower impact on SoR. Contrarily, alterations of aspects of reality for which we have previously experienced little variation should have higher impact on our SoR. For example, changing sensorimotor contingencies (e.g., a delay between actions and their consequences) or the laws of nature (e.g., changes in gravity) for which we normally have no experience (i.e. narrow

likelihood distribution)[34–36]. Therefore, we suggest that SoR is grounded in a process of inference based on the learned distributions of the world.

Previous influential work on reality monitoring e.g.,[7,37–39] has focused on the ability to discriminate the origin of information in memory, discerning between perception and imagination. This line of research has concentrated on the mechanisms allowing the separation of internally and externally generated information, based on the proposition that hallucinations may be due to failures in discriminating between sources of information (i.e., perception vs. imagery or memory). Source monitoring paradigms, thus, typically compare the ability to remember the source of a stimuli (e.g., Was this word previously seen or imagined? Spoken or heard?)[40,41]. It has been suggested that this source discrimination is grounded in the assessment of cues relating to characteristics (e.g., clarity, level of detail) of the memory[37,42,43]. Other compelling work has studied the interactions between imagery and perception showing how imagination may impact the perception of real stimuli[44–46]. Lastly, a recent study utilized the motion aftereffect to probe perceptual insight by dissociating perceptions from beliefs about stimuli[47].

These insightful approaches however do not address key aspects of the visual phenomenology of hallucinations (e.g., the "bizarreness" of the experience). These experimental methods do not typically induce a feeling of strangeness which is central to understanding cases of aberrant SoR processing. Moreover, hallucinatory states are notoriously difficult to study scientifically due to their transient subjective quality[48], ineffability[49] and the fact that they typically arise in the context of psychedelic, psychiatric, medical or neurological states[50,51]. Thus, a scientific approach to studying SoR would require a replicable, quantitative approach simulating these special kinds of experience.

For this purpose, we constructed an immersive virtual reality (VR) paradigm, designed to induce *Virtual Hallucinations* (VH) simulating the phenomenology of visual hallucinations in a controlled manner. We selected several aspects of reality to be examined in this project, based on the phenomenology of distorted visual reality as found in psychiatric, neurological, medical and pharmacological states[52,9,50,53]. These

alterations of reality broadly fall into three domains: (1) Changes in 'Perception', in which the visual appearance (e.g., colors, acuity) of the scene are manipulated. Such changes in perceptual qualities are well documented in hallucinations arising from pharmacological or medical origins[53–55] and are a central feature of schizotypy[56]. (2) Laws of 'Nature', in which visual aspects of laws of nature (e.g., gravity, passage of time) are altered. Changes in the experience of natural physics are common during dreaming[57,58], and are a defining feature of psychedelic and mystical experiences[59–61]. (3) Changes of 'Self', in which the participants' sense of self is manipulated through conflicts between visual signals and self-related information (e.g., sensorimotor conflicts, changes in first person perspective)[34,35,62]. Such changes in the sense of self are a hallmark of altered reality in several neurological[62–64], psychiatric and mystical experiences[34,65–68]. Importantly, the VR environment allows us to accurately manipulate such aspects, enabling us to alter reality at qualitatively different scales. For example, we can change the participants' first-person viewpoint to be as low as the floor or as high as the ceiling, simulating dramatic changes of Self which occur in hallucinatory states (e.g., Alice in Wonderland syndrome[69]). We can also induce minute changes, barely noticeable to the participants, resembling prodromal symptoms and derealization-like states[25,26]. Critically, psychophysics was used to individually fit VH intensities, enabling us to compare the subjective experience of different magnitudes and types of VH within and between participants. We present evidence from two experiments targeting SoR, an initial exploratory experiment - Experiment 1, followed by a pre-registered replication - Experiment 2[70]. Our behavioral results are accompanied by converging evidence from physiological measures (kinematics, pupillometry and cardiac activity) demonstrating physiological responses to manipulations of SoR. Finally, we utilized a Bayesian perceptual inference model providing new insights on the perceptual decision process underlying SoR.

Pre-registered hypotheses

We pre-registered our hypotheses, study design, sampling plan, and analysis methods based on findings from Experiment 1 prior to analyzing the Experiment 2 replication dataset[70].

Hypothesis 1 (H1). Subjective judgments:

a. Subjective Sense of Reality (mean of SoR ratings) will diminish with increasing VH magnitude.

b. Changes in SoR ratings between VH magnitudes will vary depending on the domain they are associated with.

Hypothesis 2 (H2). Physiological signals:

a. Hand kinematics during the reaching action will vary between objectively Altered and Unaltered trials.

b. Hand kinematics during the reaching phase will vary between trials subjectively experienced as Altered or Unaltered.

c. Pupillometric data will vary between Altered and Unaltered trials.

d. Cardiac activity will vary between trials subjectively experienced as Strongly Altered ('Strange') or Unaltered ('Normal').

e. Cardiac activity will vary between domains in trials subjectively experienced as Strongly Altered ('Strange').

Hypothesis 3 (H3). Clinical correlations:

a. We hypothesized that there will be correlations between the slope of ratings in the Self domain and total scores of self-reported clinical measures of abnormal-self and abnormal-perceptual (see Methods-Clinical measures).

Figure 1. (A) Conceptual illustration of experiencing a strange perception. Given such an experience the mind must make a judgment regarding the veracity of the perception. **(B)** *Virtual Hallucination* (VH) paradigm used to systematically manipulate SoR in a controlled experimental setting. Participants experienced an immersive VR environment with parametrically controlled perceptual distortions while their physiological responses were recorded (kinematics, eye-tracking and ECG). Here a ripple effect is shown. **(C)** Left: The experiment targeted three domains of altered reality, each corresponding to different real-world probabilities of occurrence: Perception (sensory distortions), Self (body-related alterations), and Nature (violations of physical laws). Middle: Psychophysical modeling was applied to control for differences in salience across conditions, ensuring that distortions were matched across domains and participants based on four magnitude levels. Right: the four magnitudes per condition were then used as the stimuli for the SoR rating task, presented in random order at varying magnitudes for 400 trials. We hypothesized that the rating curves for VH from different domains would diverge due to canonically different prior sensory expectations from each. **(D)** The SoR rating task was used to quantify subjective experience. On each trial, participants viewed an altered or unaltered virtual scene and rated how "Normal" to "Strange" the experience felt on a continuous scale (100–0 respectively). This approach allowed for precise measurement of the relationship between VH magnitude, domain, and behavioral/physiological responses.

Methods

Participants

63 participants were recruited for two experiments. Experiment 1 included 34 participants, of which 2 dropped out during the experiment and 1 was excluded due to insufficient data, leaving a total of 31 (mean age 27.6 years ± 5.3 years, 14 females). Based on the effect sizes found in Experiment 1, we performed a power analysis that indicated a minimum sample size of 32 was required to replicate the effects (Supplementary "Power analysis"). For Experiment 2, we recruited 40 participants, of which 8 were excluded whose responses were not sufficiently captured by psychometric fits (see pre-registration and Supplementary "Data exclusion criteria"), resulting in a final sample of 32 (mean age 25.1 ± 4.5 years, 22 females). All participants provided written informed consent (IRB ISU202012001), were right-handed with normal or corrected to normal vision, and self-reported no substance dependence and no current or prior history of significant medical or neurological illness or use of psychiatric medication.

General procedure

We implemented a within-subject design in which all participants underwent the same protocol and structure of the experiment in both groups. Participants attended two sessions, session 1 was devoted to measuring psychophysical thresholds which were used as normalized condition levels for measuring SoR across different conditions in session 2. The SoR rating task was explained by the experimenters in detail with the aid of a scripted presentation, while keeping subjects naive to the true purpose of the experiment. In both sessions participants sat in front of a table at the center of the experiment room and wore a VR headset and an ECG harness. After the task participants filled out three self-report questionnaires.

Virtual reality hardware and environment

During the experiments participants wore an HTC Vive-pro Eye head-mounted display (HMD) with pupillometry capabilities and retractable headphones and responded using a touch-sensitive hand controller[71]. The baseline virtual environment was constructed as a furnished house interior, providing a

normative setting for subsequent reality manipulation using VH. We enhanced the realism of the environment by matching several of the task-relevant virtual items with real ones, thus adding physical correspondence to the experience (Supplementary "Setup and calibration procedures").

Experimental Design

Inside the VR environment the participants found themselves positioned at the center of the virtual house in front of a corresponding virtual table. From their seated position they were able to look around and interact with the environment. The matching dimensions of the real and virtual tables amplified the tactile, proprioceptive and visual feedback of the participants exact position and height in VR. Their right hand and controller were represented by a virtual hand and an identical virtual controller. The virtual controller and hand provided similar feedback on the correspondence between real and virtual hand motions. To ensure participants' engagement and attention to the virtual environment on each trial participants performed a sensorimotor task in which they had to point to a virtual butterfly which could appear at several locations (Supplementary "Secondary task").

Experimental Conditions

We leveraged the properties of VR to induce specific, highly-controlled alterations in our virtual environment, resembling the phenomenology observed in various psychiatric and neurological conditions[8,62,72]. We focused on three theoretical domains of experience: Nature, Self, and Perception. In the Nature domain, we manipulated the physics of how objects move in the virtual environment by raising or lowering the gravitational constant and the rate of motion, resulting in 'heavy' and 'light' gravity VH, and 'fast', a time perception VH. These VH made objects appear to move at the wrong speed or fall in unexpected ways. In the Self domain, we manipulated the participants' first-person perspective (1PP) by raising or lowering the camera height (i.e., their personal vantage point) in the 'grow' and 'shrink' VH, and their sense of embodiment and agency by creating a lag between their real and virtual hand movements in the 'delay' VH. In the Perception domain, we manipulated the purely visual aspects of the environment by

using filters that increased or decreased color saturation in the 'saturated' and 'unsaturated' color VH, or altered the texture of the scene in the 'ripple' VH (a rippling effect that distorts peripheral vision). A full description of VH implementations in Unity is detailed in Supplementary table 1.

Experimental Procedures

Psychophysics experiment

To ensure valid comparisons across VH conditions and between participants, we calibrated stimulus intensities individually using a two-alternative-forced-choice (2AFC) 1-up\1-down descending staircase procedure (see Fig. 1C Middle and Fig. S3). The order of conditions was randomized between participants. Participants viewed two sequential phases on each trial, one containing a VH, and reported whether they detected a difference. Stimulus intensity was adjusted based on detection performance until meeting acceptable fit criteria (see Supplementary "Detailed staircase procedure" for detailed methodology).

Instructions and training

Participants were instructed on the staircase procedure, VR equipment and environment before entering VR. Participants were then introduced to the VR equipment, and environment to familiarize themselves with the unaltered virtual environment. Once in VR participants were instructed again in the psychophysical staircase task. Participants trained before the task with a structured staircase as an example. On each trial, participants were immersed in the virtual environment while performing the secondary task for two phases (5.2 s each), one of which contained a VH. After viewing both phases the participant was required to judge if both phases felt the 'same' (press right) or 'different' (press left).

Psychophysical curve estimation

We fit the psychometric curves using the Matlab Psignifit toolbox[73], and the goodness-of-fit was assessed using McFadden's pseudo-R^2 [74] (Supplementary "Psychophysical curve estimation"). We extracted four detection threshold values for each VH from the psychometric curves and used them to simulate changing

magnitudes of alteration for the SoR rating task based on their predicted detection probability: 25%, 50% (the Just Noticeable Difference or JND), 75%, and 95% (Fig. 1C). Unaltered trials were defined as having '0' magnitude, which served as our baseline condition.

SoR rating task

In the SoR rating task (Fig. 1D), participants were exposed to 400 trials of pseudo-randomly ordered conditions, 72 of which were baseline trials that did not include a VH (magnitude 0). Each trial included a single appearance of the virtual environment for 5.2 seconds, during which participants performed the secondary task as instructed and observed the state of the environment. At the end of each trial the virtual environment disappeared followed by a gray question screen during which the participants were instructed to rate their experience in the virtual room (compared to the baseline condition) on a scale from Normal to Strange (100-0 respectively) using the VR controller.

Instructions & training

Participants were instructed on the task, then entered the environment to refamiliarize themselves with the unaltered virtual environment. Training consisted first of the participants performing 9 trials using only 4th magnitude stimuli for each condition. The rest of training consisted of an additional 18 randomly ordered trials simulating the actual task. Following training, participants immediately commenced the SoR rating task. An intermittent break was offered half way through to avoid fatigue. To avoid VR motion sickness, participants were instructed to report any discomfort and could stop the experiment at any stage.

Clinical measures

After the experiment, participants completed three self-report questionnaires measuring symptoms related to SoR such as psychosis and altered perception: the Cardiff Anomalous Perceptions Scale (CAPS)[75], the Prodromal Questionnaire Brief Version (PQ-B)[76], and the Inventory of Psychotic-Like Anomalous Self-Experiences (IPASE)[77].

Data analysis

All statistical analyses were performed in R 4.4.1 and Python 3.9. Non-parametric tests (Wilcoxon signed-rank test, Spearman correlations and Welch's tests) were employed where normality assumptions were violated. Following standard procedures, we computed effect sizes for Wilcoxon signed-rank tests as $r = Z/\sqrt{N}$, where N is the number of participants. For mixed effects analyses, we used the lme4 package with effects reported using likelihood ratio tests (χ^2). For experiment 2 all statistical tests were one-sided based on our pre-registered hypotheses unless otherwise specified.

To assess physiological markers of reality deviation, we implemented machine learning classifications using logistic regression (scikit-learn, imblearn libraries). To account for data variability, we created 1000 different test-train split stratifications per subject. For imbalanced datasets, we applied oversampling using scikit-learns standard procedures. Model parameters were set to L2 penalty, C=10, with maximum iterations of 1000. Classification performance was evaluated using area under the curve (AUC) and F1 scores averaged across iterations. For cross-experiment validation, we created a composite model by averaging coefficients and intercepts across individual models from Experiment 1, which was then tested on Experiment 2 data. Data visualization and statistical reporting follow standard conventions, with error bars representing 95% confidence intervals.

The effects of VH magnitude and domain on Sense of Reality (SoR) ratings were analyzed using a linear mixed effects model. Trials were categorized by objective presence of VH ('Altered' vs. 'Unaltered'), by objective strength ('Unaltered' [magnitude 0], 'Weakly altered' [1st & 2nd magnitudes], and 'Strongly altered' [3rd & 4th magnitudes]), and by subjective experience ('Normal' & 'Strange', based on the 1st and 4th quartiles, respectively, of the min-max normalized ratings per participant). Domain-specific response patterns were quantified using slopes derived from the random effects of a linear mixed effects model. Finally, we examined correlations between clinical scores and SoR metrics (average ratings and slopes per

domain) using Spearman correlation, with outliers removed using the standard deviation method (mean ± 2.5 SD).

Physiological measures

Based on the evidence that VH impacted phenomenological judgments of reality, we examined whether exposure to VH also influenced physiological measures. Specifically, we investigated if being exposed to altered reality modifies kinematic, pupillometric, or cardiac responses. To facilitate this analysis across all three signals, we employed a subgroup approach. We categorized trials using subjective (Strange vs. Normal experiences) and objective measures (Strongly altered vs. Weakly altered and Unaltered), allowing us to compare physiological responses between these subgroups. This strategy enabled us to detect potential differences in physiological patterns associated with varying degrees and perceptions of reality alterations. We conducted two types of analyses for these measures: hypothesis driven statistics reported first demonstrate the strength and reliability of physiological responses, and complementary machine learning classifications reaffirm the robustness of our pre-registered analyses, and demonstrate the potential to classify and predict SoR using complex signal combinations.

Kinematics

The kinematic data of 60 participants were analyzed (n = 30 from Experiment 1 and n = 30 from Experiment 2). Three participants were excluded for having insufficient data (over 10% of trials excluded) based on our pre-registered criteria (Supplementary, "Data exclusion criteria"). Velocity measures were computed from the positional data. The hand movements were segmented into two phases: reaching and retracting. The reaching phase began with movement initiation toward the butterfly and ended when the velocity on the Y-axis reached 0 m/s. The retraction phase was determined using the same criteria.

Pupillometry

The pupil size data for 59 participants were analyzed (n = 30 from Experiment 1 and n = 29 from Experiment 2). Four participants were excluded for insufficient data (less than 8 trials per condition) based on our pre-

registered exclusion criteria. We chose our main time window of interest between the 4th and 5th seconds after trial initiation[78–80]. This allowed us to capture the delayed peak in pupil responses while ensuring pupil stabilization following luminance transitions between the virtual environment and the gray question screen.

Heart rate

The heart rate data from 47 participants were analyzed (n = 24 from Experiment 1 and n = 23 from Experiment 2). We excluded 16 participants from analysis due to insufficient data (over 100 excluded trials) based on our pre-registered criteria. QRS complexes were detected and R-R intervals were used to calculate heart rate. Heart rate trajectory was down-sampled and averaged into 12 epochs of 430 ms. Heart rate change trajectory was computed as the difference of each epoch from the average of heart rate recorded in the 500 ms pre-trial period.

In the exploratory analysis of the cardiac data from Experiment 1, we observed distinct patterns in the averaged heart rate trajectories for trials with different subjective experiences (trials defined as Normal or Strange based on the 1st and 4th quartiles of min-max normalized ratings) and for trials from different domains. These observations led us to focus our analysis on time-locked heart rate responses to trial content using the averaged heart rate trajectory as our primary measure. To account for both the within-subject repeated measures nature of our data and the between-subject variability in heart rate responses, we analyzed the data using mixed effects models[70].

Computational model

As introduced earlier, SoR can be understood as a probabilistic inference process, where sensory inputs are evaluated against prior knowledge of the world. As a result, alterations in aspects of reality which we normally experience as inflexible such as the force of gravity should have a stronger impact on SoR (i.e., narrow likelihood distribution) than those we frequently experience in varied contexts such as color perception (i.e., wide likelihood distribution). To formulate this process, we modified a Bayesian causal inference model adapted from Chancel, Ehrsson & Ma (2022)[30,47,81–85], commonly used in multi-sensory

perceptual decision tasks. The model contains three key variables: the category C, the stimulus x_{trial}, and σ the sensory noise assumed by the observer. The category of the stimulus has a value of $C = Real$ if x_{trial} was Real or $C = Unreal$ if x_{trial} was Unreal. The prior probability of an event being judged as real before any sensory stimulation is $p(C = Real) = p_{real}$

On each trial the model observer receives a stimulus x_{trial} and infers its category C by computing the posterior probabilities $p(C = Real| x_{trial})$ and $p(C = Unreal| x_{trial})$. The comparison of the two posteriors is expressed as the log posterior ratio d. The posterior ratio d is compared to 0 when there is equal prior probability of C being real and unreal. In turn, the decision rule states the observer would infer x_{trial} is real if d is positive.

$$d = \log\left(\frac{p(C = Real|x_{trial})}{p(C = Unreal|x_{trial})}\right)$$

Importantly, this decision is not based solely on the posterior for reality but rather on the relative comparison of both posteriors. While the posterior for reality, $p(C = Real| x_{trial})$, is shaped by prior expectations and sensory likelihoods, the posterior for unreality, $p(C = Unreal| x_{trial})$, represents a more diffuse, non-specific belief that any stimulus—especially those at extreme values—could be unreal. This means that even familiar stimuli, such as a blue sky at noon, could theoretically be explained by an unreal experience (e.g., hallucination, simulation), but in most cases, the posterior probability for the real explanation far exceeds that of unreal.

Next, we rewrite the log ratio d as a combination of likelihoods which depend on the combination of the sensory noise σ (assumed by the participant), the signal noise σ_s (the standard deviation of the real signal values), and the posteriors p_{real} and p_{unreal}:

$$d = \log\left(\frac{p_{real}}{p_{unreal}}\right) + \log\left(\frac{N(x_{trial} \mid 0, \sigma^2 + \sigma_s^2)}{N(x_{trial} \mid 0, \sigma^2)}\right)$$

$$= log\left(\frac{p_{real}}{p_{unreal}}\right) + \frac{1}{2}log\left(\frac{\sigma^2 + \sigma_s^2}{\sigma^2}\right) - \frac{x_{trial}^2}{2}\left(\frac{1}{\sigma^2} - \frac{1}{\sigma^2 + \sigma_s^2}\right)$$

This expression can be reformulated as a threshold on sensory input k, which represents the point where the posterior for real and unreal are equal ($d = 0$).

$$\frac{x_{trial}^2}{2}\left(\frac{1}{\sigma^2} - \frac{1}{\sigma^2 + \sigma_s^2}\right) < log\left(\frac{p_{real}}{p_{unreal}}\right) + \frac{1}{2}log\left(\frac{\sigma^2 + \sigma_s^2}{\sigma^2}\right)$$

This allows a direct evaluation of d given x_{trial} and the decision criterion k:

$$x_{trial}^2 < \frac{\sigma^2(\sigma^2 + \sigma_s^2)}{\sigma_s^2}\left(2log\left(\frac{p_{real}}{p_{unreal}}\right) + log\left(\frac{\sigma^2 + \sigma_s^2}{\sigma^2}\right)\right)$$

$$k = \sqrt{\frac{\sigma^2(\sigma^2 + \sigma_s^2)}{\sigma_s^2}\left(2log\left(\frac{p_{real}}{p_{unreal}}\right) + log\left(\frac{\sigma^2 + \sigma_s^2}{\sigma^2}\right)\right)}$$

Thus, the decision rule $d > 0$ is equivalent to $\rightarrow |x| < k$.

Figure 2. Computational model of Sense of Reality (SoR). A conceptual illustration of the proposed inference-based model of SoR, where stimuli are classified as Real or Unreal based on the posterior probability ratios between the two possibilities. The model incorporates prior expectations and sensory noise, leading to dynamic decision thresholds. **(A)** When experiencing an unreal

event, the decision to classify it as real or unreal is determined by the log-posterior ratio of the two competing posteriors: $p(C = Real \mid x_{trial})$ and $p(C = Unreal \mid x_{trial})$. The likelihood of real sensory alterations is modeled as a half Gaussian (light-green). The likelihood for unreal sensory alterations is modeled as a wide, shallow Gaussian (light-red), representing a generalized, non-specific belief that all experiences could be unreal, especially extreme deviations from the norm. These two distributions overlap in the region where both hypotheses assign probability to the observed sensory alteration (darker green). The observer's decision criterion k is defined as the point where the posteriors are equal (i.e., where the green and red likelihoods intersect, and the log-posterior ratio equals zero). Smaller, more frequent alterations fall to the left of this threshold and are more likely judged as real, while larger, more unusual alterations fall to the right and are more likely judged as unreal. **(B)** Illustration of how prior probability (p_{real}) and sensory noise (σ) shape the decision rule (k). Left: The effect of varying the prior probability of reality p_{real}, with values of 0.9 (blue), 0.8 (red), and 0.6 (green). Higher priors for reality lead to more conservative SoR judgments, especially for weaker alterations. Right: The effect of varying sensory noise (σ), with values of 0.7 (blue), 0.6 (red), and 0.4 (green). Increased sensory noise broadens the likelihood function, shifting the decision criterion k. **(C)** Model-predicted effects of p_{real} and σ on SoR judgments. Left: Simulated changes in p_{real} influence the expected probability for judging an event as real. Right: Simulated changes in σ influence the width of the distribution, demonstrating how higher expected noise increases tolerance for deviations from expected reality.

We fitted this model to participants' SoR ratings across different domains and magnitudes. Since all participants experienced stimuli at five magnitudes corresponding to their individual psychophysical performances, the model accounted for these variations. In this model the response depends on the prior probability that x_{trial} is real (p_{real}) and the sensory noise in each trial (σ). The model has two more free parameters: λ would be the lapse rate accounting for random guesses or unintended responses and we allowed an individual-level bias in using the SoR scale, by including a bias parameter that accounts for differences in average ratings between participants.

$$p(Real \mid x_i) = Bias + 0.5\lambda + (1 - \lambda)(\Phi(k; x_i, \sigma^2) - \Phi(-k; x_i, \sigma^2))$$

We then fitted the model to best capture the participants' ratings, independently for each participant and each domain, using a root mean squared distance (RMSD) as a measure of fitting error:

$$RMSD_{(\text{Participant,Domain})} = \sqrt{2 \sum_{i=1}^{N} (SoR_i - p(Real \mid x_i))^2}$$

The *optim* function was used to find the set of parameters $\{\lambda, Bias, p_{real}, \sigma\}$ that minimized the $RMSD$ error function for each participant in each domain. The output of the model is the probability of the observer deciding that x_{trial} is real.

Pre-registered Hypotheses: Notation Guide

The relationships of reported results to pre-registered hypotheses are indicated as follows:

† Indicates a fully confirmed pre-registered hypothesis; ‡ indicates a partially confirmed pre-registered hypothesis; × indicates an unconfirmed pre-registered hypothesis. Where applicable, the specific hypothesis number is referenced (e.g., [H1.a †]).

Results

Unless otherwise specified, all analyses reported below were pre-registered. Unregistered exploratory analyses are introduced separately toward the end of the results section (see "Classifying SoR from physiological signals" and "SoR ratings fit to a Bayesian causal inference model").

SoR varies across domains and magnitudes. Analysis using mixed-effect models on the SoR ratings revealed a significant main effect of magnitude on ratings in Experiment 1 ($\chi^2(4) = 412$, $p < .0001$, marginal $R^2 = 0.36$, conditional $R^2 = 0.77$) and Experiment 2 ($\chi^2(4) = 439$, $p < .0001$, marginal $R^2 = 0.33$, conditional $R^2 = 0.80$), suggesting that SoR ratings decreased along with sensory uncertainty, see Table S4 and Fig. S5A [H1.a †].

The main effect of domain was also significant in Experiment 1 ($\chi^2(2) = 17$, $p = .0002$, marginal $R^2 = 0.05$, conditional $R^2 = 0.3$; see Fig. 3), reflecting a tendency to rate Nature VH as most unreal and Perception alteration as least unreal (Nature: 61.4, Self: 67.2, Perception: 72.6). The main effect of domain did not replicate in Experiment 2 ($\chi^2(2) = 3.6$, $p = 0.15$, marginal $R^2 = 0.005$, conditional $R^2 = 0.4$). However, a significant interaction effect between magnitude and domain was found both in Experiment 1 ($\chi^2(8) = 47.4$, $p < .001$, marginal $R^2 = 0.44$, conditional $R^2 = 0.79$) and Experiment 2 ($\chi^2(8) = 19.2$, $p = .013$, marginal $R^2 = 0.35$, conditional $R^2 = 0.81$) [H1.b †]. Together these results show that SoR ratings varied across VH

magnitudes differently depending on the domain of the VH. See Supplementary "Mixed-models" for a detailed report of model comparisons.

Total questionnaire scores for abnormal perceptual symptoms were low to non-clinical in our sample[75–77] : CAPS (5.2 ± 4.8); PQ-B (3.8 ± 3.4); IPASE (78.4 ± 19.8). There were no significant correlations between the present behavioral measures of SoR and self-reports of abnormal perception: the mean rating per domain (Spearman's ρ ranged from -0.2 to 0.03, all $p \geq 0.14$) and the random slope per domain (using Nature as the reference level of the model): Self slope ($0 \leq \rho \leq 0.12$, all $p \geq 0.38$), [H3.a ×]; Perception slope ($0 \leq \rho \leq 0.20$, all $p \geq 0.13$) (Fig. S9).

Figure 3. Sense of Reality varies across domains. Group-average SoR ratings as a function of VH magnitude, shown separately for Experiment 1 (Left: $n = 31$), Experiment 2 (Middle: $n = 32$), and the combined dataset (Right: $N = 63$). Colored lines represent domain effects, Perception (blue), Self (red), and Nature (green)**.** Individual data points represent each participant's mean SoR rating per condition, error bars are bootstrapped 95% CI for each magnitude × domain. Linear mixed effects modeling revealed significant main effects of magnitude and a significant magnitude × domain interaction effect in both experiments and the combined data.

Physiological markers of altered reality. To understand physiological responses to altered reality, we analyzed three physiological markers – hand kinematics, pupil dilation, and heart rate - comparing them by VH presence (i.e., Altered vs. Unaltered trials) and subjective experience (i.e., Strange vs. Normal trials) (see "Data analysis"). Analysis of kinematic data from Experiment 1 revealed a difference in movements during VH. Specifically, mean hand velocity during the reaching phase was higher in Altered compared to Unaltered trials (0.47 ± 0.16 m/s vs. 0.44 ± 0.17 m/s; 13.48% increase; W(29) = 23, $p < .0001$, Wilcoxon's $r = 0.71$) and in trials subjectively experienced as Strange compared to trials experienced as Normal (0.48 ± 0.16 m/s vs. 0.45 ± 0.17 m/s; 17.08% increase; W(29) = 33, $p < .0001$, Wilcoxon's $r = 0.71$). These effects were replicated in Experiment 2 (Altered vs. Unaltered: 0.39 ± 0.14 m/s vs. 0.37 ± 0.15 m/s; 8.85% increase; t(29) = -3.59, $p < .0001$, Cohen's $d = 0.65$) [H2.a †]; (Strange vs. Normal: 0.39 ± 0.14 m/s vs. 0.37 ± 0.15 m/s; 13.82% increase; t(29) = -5.11, $p < .0001$, Cohen's $d = 0.93$) [H2.b †]. Thus, objective and subjective aspects of VH were both associated with significant changes in sensorimotor behavior (Fig. 4A & Fig. S6).

Differences in pupillometry were also found during VH. Average pupil size increased in the 4000-5000 ms window of Experiment 1 (Δ 0.57 ± 0.18 mm; t(29) = 4.15, $p < .0001$, Cohen's $d = 3.1$) and replicated in Experiment 2 (Δ 0.67 ± 0.22 mm; t(28) = 3.78, $p < .0001$, Cohen's $d = 3$) [H2.c †] (Fig. 4C). Additional analysis revealed the effect remained significant when comparing over the entire trial duration (0-5000 ms) in both experiments: Experiment 1 (Δ 0.36 ± 0.12 mm; t(29) = 3.18, $p < .001$, Cohen's $d = 3$); Experiment 2 (Δ 0.41 ± 0.16 mm; t(28) = 2.55, $p < .001$, Cohen's $d = 2.5$), demonstrating a robust association between pupil dilation and VH presence.

Heart rate was found to decelerate during trials experienced as Strange compared to trials experienced as Normal in Experiment 1 ($\chi^2(1) = 4.38$, $p = 0.036$, Cohen's $d = 0.47$) and Experiment 2 ($\chi^2(1) = 15.3$, $p < .0001$, Cohen's $d = 0.81$) [H2.d †] (Fig. 4E). Interestingly, in Strongly altered trials heart rate deceleration showed a significant domain effect in Experiment 1 ($\chi^2(2) = 14.63$, $p < .001$, Cohen's $d = 0.78$), which was

most pronounced in Nature (β = -0.38), followed by Self (β = -0.17), and Perception (Fig. S7). However, this effect did not replicate in Experiment 2 ($\chi^2(2)$ = 5.82 p = 0.054, Cohen's d = 0.49) [H2.e ×]. Leading us to conclude that heart rate deceleration was only consistently associated with the subjective strangeness of the trial, but not with the domain of the VH.

Figure 4. Physiological markers of altered reality.
(**A, C, E**) Individual changes in physiological responses for each participant, based on paired comparisons from the pre-registered analysis plan. Each bar represents the relative change per subject in kinematic (**A**), pupillometric (**C**), and cardiac (**E**) measures between Altered and Unaltered trials. (**B, D, F**) Single-subject classification performance (AUC) for predicting Altered vs. Unaltered trials using kinematic (**B**), pupillometric (**D**), and cardiac (**F**) measures. Each dot represents an individual participant; error bars show standard error, and the dashed line indicates chance level (AUC = 0.50). Classification was significantly above chance in both experiments (Exp 1: AUC = 0.58; Exp 2: AUC = 0.55–0.59, all p < .0001). (**G**) Multi-modal classification performance, integrating different physiological measures. Icons indicate which physiological modalities were combined. Dual-modality combinations improved classification accuracy, reaching AUC ≈ 0.62 (Exp 1) and 0.60–0.61 (Exp 2), while three-signal combinations achieved the highest performance (AUC = 0.65 and 0.64 for Exp 1 and Exp 2, respectively).

Classifying SoR from physiological signals. The following analyses were conducted as an exploratory extension beyond our pre-registered analyses to assess whether physiological measures could predict SoR ratings. We evaluated the independent and combined potentials of physiological signals to predict fundamental changes in SoR such as VH presence (Altered vs. Unaltered trials) and subjective experience of VH (Strange vs. Normal trials) with a multi-faceted classification approach using linear regression with Area Under the Curve (AUC) as our performance metric. Firstly, Our analysis revealed that each physiological signal independently contained information allowing for above chance classification of Altered vs. Unaltered trials in Experiment 1 (Kinematics: AUC = 0.58 ± 0.03, t(29) = 12.65, $p < .0001$, Cohen's d = 2.667 ; Heart rate: AUC = 0.58 ± 0.02, t(24) = 19.67, $p < .0001$, Cohen's d = 4; Pupil size: AUC = 0.58 ± 0.04, t(29) = 10.40, $p < .0001$, Cohen's d = 2), and in Experiment 2 (Kinematics : AUC = 0.57 ± 0.03, t(29) = 14.83, $p < .0001$, Cohen's d = 2.333 ; Heart rate: AUC = 0.59 ± 0.02, t(23) = 16.65, $p < .0001$, Cohen's d = 4.5; Pupil size: AUC = 0.55 ± 0.04, t(28) = 6.95, $p < .0001$, Cohen's d = 1.25) (Fig. 4B, D & F).

Moreover, we found that apart from their independent contributions, all signal combinations improved classification success in Experiment 1 (Kinematics & Heart rate: AUC = 0.62 ± 0.03, t(24) = 18.26, $p < .001$, Cohen's d = 4; Kinematics & Pupil size: AUC = 0.62 ± 0.04, t(29) = 14.60, $p < .001$, Cohen's d = 3; Pupil size & Heart rate: AUC = 0.62 ± 0.04, t(24) = 15.06, $p < .001$, Cohen's d = 3; Pupil size, Heart rate & Kinematics: AUC = 0.65 ± 0.04, t(24) = 17.60, $p < .001$, Cohen's d = 3.75), and in Experiment 2 (Kinematics & Heart rate: AUC = 0.61 ± 0.02, t(23) = 21.59, $p < .001$, Cohen's d = 5.5; Kinematics & Pupil size: AUC = 0.60 ± 0.05, t(28) = 9.86, $p < .001$, Cohen's d = 2; Pupil size and Heart rate: AUC = 0.60 ± 0.03, t (20) = 15.00, $p < .001$, Cohen's d = 3.3; Pupil size, Heart rate & Kinematics: AUC = 0.64 ± 0.05, t(20) = 12.17, $p < .001$, Cohen's d = 2.8) (Fig. 4G). Notably, a cross-subject model trained on pupil data from Experiment 1 successfully classified VH presence in Experiment 2 (AUC = 0.53 ± 0.03, t(28) = 4.99, $p < .001$, Cohen's d = 1) (Fig. S8C). These findings demonstrate that physiological signals contain quantifiable objective and subjective information about Sense of Reality, with each signal contributing

independently and their different combinations improving predictive accuracy. The generalizability of changes in pupil size across experiments further supports its potential as a robust marker for experiences of altered reality.

SoR ratings fit to a Bayesian causal inference model. The following computational model analysis was not pre-registered but was conducted post hoc to explore the decision mechanisms underlying SoR ratings. Our model assumes the observer infers the category C of x_{trial} to decide if it is real or unreal. The model depends on the prior probability for an event being real and the sensory noise in each trial. The model has four free parameters: p_{real} the prior probability of x_{trial} being real, independent of any stimuli, σ the sensory noise, λ the lapse rate for random guessing or mistakes, and a bias parameter accounting for individual differences in ratings. The model fits the data well, effectively simulating the observed domain differences (Fig. 5A-B), supporting our hypothesis that SoR is a process of inference.

Each parameter was analyzed using a separate linear mixed-effects model with domain as a fixed effect and participant as random intercept. The analyses revealed a significant main effect of domain for p_{real} (F(2,124) = 17.9, p < .0001, marginal R^2 = 0.13, conditional R^2 = 0.3); σ (F(2,124) = 5.3, p < .01, marginal R^2 = 0.04, conditional R^2 = 0.29) and k (F(2,124) = 4.9, p < .01, marginal R^2 = 0.03, conditional R^2 = 0.3), indicating that domain systematically influenced prior probability, sensory noise and decision criterion (see Fig. 5C-E). Post-hoc pairwise comparisons using estimated marginal means with Bonferroni correction revealed that Nature had the lowest p_{real} (0.58), significantly different from Perception and Self (both ~ 0.7; t(124) = -5, p < .0001, Cohen's d = -0.9), suggesting that participants were biased to assume the least prior probability for Nature VH being real. Similarly, Nature had the lowest σ, with a significant difference observed between Nature and Perception (t(124) = -3.1, p < 0.01, Cohen's d = -0.56), reflecting a narrower likelihood distribution for Nature and a wider one for Perception (Nature: 0.36, Self: 0.38, Perception: 0.43). Finally, Nature exhibited the strictest decision criterion k (0.61), significantly lower than Self (0.98; t(124) = -2.6 , p = 0.03, Cohen's d = -0.46) and Perception (1.0; t(124) = -2.8, p = .016, Cohen's d = -0.5),

reinforcing that domain differences in SoR judgments are influenced by priors, likelihood distributions, and decision criteria.

Figure 5. A Bayesian causal inference model relates changes in SoR to domain-specific prior expectations. **(A)** Model predictions (colored points) overlaid on observed group mean SoR ratings (colored lines) as a function of VH magnitude. **(B)** Observed SoR ratings per domain (bars) alongside model-predicted ratings (shaded regions). **(C–E)** Comparison of key model parameters across domains. Colored points represent individual model predictions, boxplots indicate the 1st–3rd quartile range, and horizontal lines show the median value for each domain. **(C)** Prior probability (p_{real}) was lowest for Nature (0.58), followed by Perception (0.7) and Self (0.71), indicating that participants assumed the least prior probability of reality for alterations in Nature. **(D)** Sensory noise (σ) was lowest for Nature (0.36), then Self (0.38), and highest for Perception (0.43), reflecting narrower likelihood distributions for physical law violations and greater variability in perceptual distortions. **(E)** Decision criterion (k) was strictest for Nature (0.61), followed by Self (0.98) and Perception (1.00), suggesting that even minor deviations in natural laws triggered lower SoR ratings.

Discussion

In this study, we employed a novel, ecological, immersive virtual reality paradigm designed to simulate the phenomenology of visual hallucinations allowing to systematically investigate the Sense of Reality (SoR). Our primary findings reveal that: (1) VH from different domains—Nature, Self, and Perception—have distinct and replicable impact on reality judgments, with nature-related alterations being perceived as the least real, followed by those related to the Self and Perception domains. (2) Experiencing VH was associated with consistent physiological changes in hand kinematics, pupil size, and cardiac activity. (3) Finally, these findings were supported by a perceptual inference model, which successfully captured the domain-specific effects by accounting for differences in prior expectations such as the variability of sensory evidence.

Reality judgments differ across domains

We used a novel approach in which we simulated in virtual reality the reported phenomenology of nine different types of hallucinations to study their effects in a controlled manner. Utilizing psychophysical staircases, we equated the subjective saliency of these different VH and demonstrated that despite equated saliency, SoR ratings significantly differ across domains (Perception, Self, Nature), with each domain showing distinct effects on participants' reality judgments. Notably, alterations in the Nature domain were experienced as the least real, followed by alterations in the Self and Perception domains. This finding aligns with predictive processing theories[29,33,14], which suggest that violations of stable environmental regularities (such as those pertaining to the laws of nature) would lead to a larger disruption in SoR. For example, changes in the perceptual domain such as deviations in color saturation (reported in depression [86] and psychedelic[87] hallucinations) effected reality judgments less than deviations in gravity, even when controlling for salience at the single subject level. We propose that SoR judgments are based on accumulated prior experience of the observer, thus domains for which a wider range of experiences (wide likelihood distributions) are available will allow more flexibility before inducing a feeling of disturbed reality.

Physiological correlates of altered SoR

Our study provides robust evidence that exposure to VH not only affects subjective reality judgments but also induces changes in physiological responses. Specifically, we observed significant and replicable modulations in hand kinematics, cardiac activity, and pupil size during VH trials. These findings underscore the interplay between perceptual processes and the body's physiological state, indicating the integration between predictive mechanisms and sensorimotor and autonomic functions.

When observing VH hand movement velocity was higher, allowing to classify trial by trial exposure to altered reality states well above chance level. Modulation of sensorimotor processes are well documented in psychiatric and neurological disorders involving hallucinatory experiences such as Parkinson's disease and schizophrenia[88–92]. We suggest that the violated sensory expectations elicited by the VH may modify kinematic behaviors, manifested in a more automatic and rapid reaching movement, possibly related to affective sensations of uncanniness[93,94]. Indeed, hand kinematics allowed to classify not only objective exposure to VH but also subjective feelings of disturbed reality (Fig. S6). The ability to detect these subtle changes in hand movement provides a valuable objective marker for the presence of altered reality states, offering a non-invasive method to quantify the impact of hallucinations on motor function. Cardiac responses also showed significant deceleration during trials experienced by the participants as strongly deviating from reality (Strange trials). Previous research has linked cardiac responses to changes in the sense of self and consciousness[95–97], psychedelic hallucinations[98] and perceptual inference[99–101]. Specifically, heart rate deceleration predicted subjective feelings of strangeness and may be a parasympathetic response triggered by the brain's recognition of an incongruence between sensory input and internal predictions[101,102]. Finally, our results indicated a clear and replicable response of pupil dilation to VH. Pupillometry, correlated with norepinephrine (NE) release[103–105], serves as a direct and implicit measure of various internal cognitive processes[106,107], including prediction errors[108,109]. In the context of our study, the observed pupil dilation may be a response to the cognitive effort required to process unexpected or conflicting sensory information. Critically, pupillometry allowed not only to classify altered

trials within participants (Fig. 4D), but also allowed cross participant classification indicating a characteristic and generalizable pupil response to alterations of reality (Fig. S8C). Moreover, pupil dilation scaled between different magnitudes of alteration suggesting that it tracks the magnitude of violated predictions (Fig. S8A-B). This marker thus offers a real-time, non-invasive measure of the processing of perceptual reality, which can be particularly valuable in detecting and assessing the severity of hallucinations.

The identification of these physiological markers presents a special opportunity for advancing the study of hallucinations and related perceptual disturbances. Traditionally, the assessment of hallucinations has relied heavily on self-report[72], which are inherently subjective. The introduction of putative objective physiological measures offers a complementary approach, allowing for the continuous, real-time monitoring of hallucination-related states without relying solely on patient self-report. This objective approach is particularly advantageous in clinical settings where patients may have limited insight into their hallucinations or may be unable or unwilling to communicate their experiences. By providing a quantifiable and reproducible method to detect and track hallucinations, these physiological markers could be integrated into clinical assessments, enabling more accurate diagnosis and monitoring of conditions such as schizophrenia, Parkinson's disease, and other disorders characterized by perceptual disturbances.

A perceptual inference model of SoR

Our computational model demonstrates how prior expectations—such as the prior probability assigned to an event and the assumed width of the likelihood distribution—systematically influence the decision rule for reality judgments. The main strength of the model is its ability to explain domain differences. Given that saliency was normalized across conditions, one might not necessarily expect ratings to differ between domains. Yet, the model predicts their presence in a manner consistent with our conceptual framework and

in line with previous findings[110,111], reinforcing the idea that sense of reality depends on prior expectations corresponding to the distribution of environmental regularities.

While the model provides a compelling fit to the data, we do not claim it to be the definitive or most complete model for SoR. Rather, it serves as a preliminary conceptual model, demonstrating how domain differences in SoR judgments can emerge from variations in priors and likelihoods. Future modeling efforts should incorporate additional factors such as learning, attention, and the influence of clinical symptoms[112], which will likely contribute to a more nuanced and dynamic understanding of the sense of reality.

Despite its conceptual nature, this framework raises interesting questions: How are perceptual reality priors initially formed? do they update over time? Furthermore, these findings may have implications for the study of clinical conditions where SoR is impaired, such as psychosis, which has been closely linked to abnormal predictive processing. Future studies could delve into the relations between clinical states in which SoR is compromised such as psychosis and derealization [66,72], expanding the current computational model to better understand these states. Moreover, our experimental approach significantly facilitates the capacity for neuroimaging of hallucination like states. Further studies using fMRI and EEG could allow to understand the neural systems underlying SoR. Finally, while disrupted SoR is associated with psychopathology in adults, young children experience reality more flexibly[113–115]. This developmental perspective raises fundamental questions about how our model of reality is shaped over time, informing our understanding of both typical development and abnormal SoR.

Limitations

A potential limitation of this study is that participants may have experienced the environment as Virtual thus treating all experiences as "Unreal". Indeed, participants were fully aware that the VH were not real hallucinations. However, the immersive nature of VR experiences allow 'bottom-up" simulations stimulating realistic responses [116–118] despite knowledge of the virtual medium. Moreover, the large changes in subjective judgments of reality (effect sizes $R^2 \geq 0.33$) and physiological responses indicate that our

experimental approach elicited simulations of unusual experiences. Another limitation is that the current study used only two sensory modalities to test SoR (visual and motor), but there are many other relevant scenarios to simulate, such as auditory hallucinations, feelings of presence and multimodal hallucinations which are common symptoms of schizophrenia[119–121]. Finally, our neurotypical sample had low scores in all clinical measures, reducing our ability to detect possible relations between symptoms and SoR ratings.

Conclusions

In conclusion, we developed a novel experimental approach allowing to rigorously test the sense of reality by simulating hallucinations. Our results demonstrate how sense of reality judgments are dependent on internal models of the world, and how experiences deviating from these models induce changes in phenomenology and physiological markers. This framework significantly advances our understanding of hallucinatory experiences and may advance tools for assessing and treating clinical symptoms of altered sense of reality.

Author contributions

R.S conceived the original ideas and methodology. A.Z assisted with psychophysical modeling. Y.Z built the virtual environment and virtual hallucinations. R.S, G.D, P.B.T developed the tasks and experimental conditions. G.D, P.B.T, O.H, A.B and N.G performed the experiments and did the final analysis and visualizations. U.H adapted and developed the computational model and contributed to its interpretation. R.S and G.D discussed and interpreted the results and wrote the first draft and final version of the manuscript. R.S acquired funding.

Acknowledgments

This research project has received funding from an ERC starting grant (UNREAL- 949010) to Roy Salomon.

Conflict of interest

The authors declare no conflict of interest.

References

1.	Marr, D. & Poggio, T. A computational theory of human stereo vision. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **204**, 301–328 (1979).

2.	Palmer, S. E. *Vision Science: Photons to Phenomenology*. (MIT press, 1999).

3.	Hoffman, D. D. The interface theory of perception. *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* **2**, 1–24 (2018).

4.	Lau, H. Consciousness, Metacognition, & Perceptual Reality Monitoring. (2019) doi:10.31234/osf.io/ckbyf.

5.	American Psychiatric Association. *DSM 5*. (American Psychiatric Association, 2013).

6.	Barnes, J., Boubert, L., Harris, J., Lee, A. & David, A. S. Reality monitoring and visual hallucinations in Parkinson's disease. *Neuropsychologia* **41**, 565–574 (2003).

7.	Bentall, R. P., Baker, G. A. & Havers, S. Reality monitoring and psychotic hallucinations. *British Journal of Clinical Psychology* **30**, 213–222 (1991).

8.	Ohayon, M. M. Prevalence of hallucinations and their pathological associations in the general population. *Psychiatry Research* **97**, 153–164 (2000).

9.	Waters, F. *et al.* Visual Hallucinations in the Psychosis Spectrum and Comparative Information From Neurodegenerative Disorders and Eye Disease. *Schizophr Bull* **40**, S233–S245 (2014).

10.	Fletcher, P. C. & Frith, C. D. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat Rev Neurosci* **10**, 48–58 (2009).

11.	Corlett, P. R. *et al.* Hallucinations and Strong Priors. *Trends in Cognitive Sciences* **23**, 114–127 (2019).

12.	Griffin, J. D. & Fletcher, P. C. Predictive Processing, Source Monitoring, and Psychosis. *Annu Rev Clin Psychol* **13**, 265–289 (2017).

13.	Suzuki, K., Roseboom, W., Schwartzman, D. J. & Seth, A. K. A Deep-Dream Virtual Reality Platform for Studying Altered Perceptual Phenomenology. *Scientific Reports* **7**, 1–11 (2017).

14.	Corlett, P., Frith, C. D. & Fletcher, P. From drugs to deprivation: a Bayesian framework for understanding models of psychosis. *Psychopharmacology* **206**, 515–530 (2009).

15.	Powers, A. R., Mathys, C. & Corlett, P. R. Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. *Science* **357**, 596–600 (2017).

16.	O'Callaghan, C. *et al.* Visual Hallucinations Are Characterized by Impaired Sensory Evidence Accumulation: Insights From Hierarchical Drift Diffusion Modeling in Parkinson's Disease. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* **2**, 680–688 (2017).

17.	Ellson, D. G. Hallucinations produced by sensory conditioning. *Journal of Experimental Psychology* **28**, 1 (1941).

18.	Frith, C. D. & Done, D. J. Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. *Psychological medicine* **19**, 359–363 (1989).

19.	Teunisse, R. J., Cruysberg, J. R., Hoefnagels, W. H. & Verbeek, A. L. Visual hallucinations in psychologically normal people: Charles Bonnet's syndrome. *The Lancet* 794–797 (1996).

20.   Perivoliotis, D. *et al.* Cognitive insight predicts favorable outcome in cognitive behavioral therapy for psychosis. *Psychosis* **2**, 23–33 (2010).

21.   Gerretsen, P. *et al.* Impaired insight into illness and cognitive insight in schizophrenia spectrum disorders: resting state functional connectivity. *Schizophrenia research* **160**, 43–50 (2014).

22.   Nair, A., Palmer, E. C., Aleman, A. & David, A. S. Relationship between cognition, clinical and cognitive insight in psychotic disorders: A review and meta-analysis. *Schizophrenia Research* **152**, 191–200 (2014).

23.   Moritz, S. & Woodward, T. S. A generalized bias against disconfirmatory evidence in schizophrenia. *Psychiatry research* **142**, 157–165 (2006).

24.   Moritz, S., Woodward, T. S. & Ruff, C. C. Source monitoring and memory confidence in schizophrenia. *Psychological medicine* **33**, 131–139 (2003).

25.   Simeon, D., Knutelska, M., Nelson, D. & Guralnik, O. Feeling Unreal: A Depersonalization Disorder Update of 117 Cases. *The Journal of Clinical Psychiatry* **64**, 990–997 (2003).

26.   Koren, D., Lacoua, L., Rothschild-Yakar, L. & Parnas, J. Disturbances of the Basic Self and Prodromal Symptoms Among Young Adolescents From the Community: A Pilot Population-Based Study. *Schizophr Bull* **42**, 1216–1224 (2016).

27.   Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* **36**, 181–204 (2013).

28.   Rao, R. P. & Ballard, D. H. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience* **2**, 79 (1999).

29.   Shams, L. & Beierholm, U. Bayesian causal inference: A unifying neuroscience theory. *Neuroscience & Biobehavioral Reviews* **137**, 104619 (2022).

30.   Körding, K. P. *et al.* Causal inference in multisensory perception. *PLoS one* **2**, e943 (2007).

31.   Ma, W. J. Bayesian decision models: A primer. *Neuron* **104**, 164–175 (2019).

32.   Körding, K. P. & Wolpert, D. M. Bayesian integration in sensorimotor learning. *Nature* **427**, 244–247 (2004).

33.   Shams, L. & Beierholm, U. R. Causal inference in perception. *Trends in cognitive sciences* **14**, 425–432 (2010).

34.   Krugwasser, A. R., Harel, E. V. & Salomon, R. The boundaries of the self: The sense of agency across different sensorimotor aspects. *Journal of Vision* **19**, 14–14 (2019).

35.   Salomon, R. *et al.* Changing motor perception by sensorimotor conflicts and body ownership. *Scientific Reports* **6**, 25847 (2016).

36.   Merfeld, D. M., Zupan, L. & Peterka, R. J. Humans use internal models to estimate gravity and linear acceleration. *Nature* **398**, 615–618 (1999).

37.   Johnson, M. K. Reality monitoring: An experimental phenomenological approach. *Journal of Experimental Psychology: General* **117**, 390 (1988).

38.   Johnson, M. K., Hashtroudi, S. & Lindsay, D. S. Source monitoring. *Psychological bulletin* **114**, 3 (1993).

39.   Johnson, M. K. & Raye, C. L. Reality monitoring. *Psychological review* **88**, 67 (1981).

40.   Vinogradov, S. *et al.* Brain activation patterns during memory of cognitive agency. *NeuroImage* **31**, 896–905 (2006).

41.   Raye, C. L. & Johnson, M. K. Reality monitoring vs. discriminating between external sources of memories. *Bulletin of the Psychonomic Society* **15**, 405–408 (1980).

42.   Garrison, J. R., Bond, R., Gibbard, E., Johnson, M. K. & Simons, J. S. Monitoring what is real: The effects of modality and action on accuracy and type of reality monitoring error. *Cortex* **87**, 108–117 (2017).

43.   Johnson, M. K., Raye, C. L., Foley, H. J. & Foley, M. A. Cognitive Operations and Decision Bias in Reality Monitoring. *The American Journal of Psychology* **94**, 37–64 (1981).

44.   Dijkstra, N. & Fleming, S. Fundamental constraints on distinguishing reality from imagination. (2021).

45.   Dijkstra, N. & Fleming, S. M. Subjective signal strength distinguishes reality from imagination. *Nat Commun* **14**, 1627 (2023).

46.   Dijkstra, N. Confidence ratings do not distinguish imagination from reality | JOV | ARVO Journals. https://jov.arvojournals.org/article.aspx?articleid=2793708 (2024).

47.   Mihali, A., Broeker, M., Ragalmuto, F. D. M. & Horga, G. Introspective inference counteracts perceptual distortion. *Nat Commun* **14**, 7826 (2023).

48.   Collerton, D. Understanding visual hallucinations: A new synthesis. *Neuroscience & Biobehavioral Reviews* **150**, 105208 (2023).

49.   Hewitt, K. The "Feeling of Knowing," the Psychedelic Sensorium, and Contemporary Neuroscience. *The Senses and Society* **6**, 177–202 (2011).

50.   Dudley, R. A comparison of visual hallucinations across disorders - PubMed. https://pubmed.ncbi.nlm.nih.gov/30579187/ (2018).

51.   Leptourgos, P. *et al.* Hallucinations Under Psychedelics and in the Schizophrenia Spectrum: An Interdisciplinary and Multiscale Comparison. *Schizophr Bull* **46**, 1396–1408 (2020).

52.   Blanke, O. & Metzinger, T. Full-body illusions and minimal phenomenal selfhood. *Trends in Cognitive Sciences* **13**, 7–13 (2009).

53.   Preller, K. H. & Vollenweider, F. X. Phenomenology, Structure, and Dynamic of Psychedelic States. in *Behavioral Neurobiology of Psychedelic Drugs* (eds. Halberstadt, A. L., Vollenweider, F. X. & Nichols, D. E.) 221–256 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2018). doi:10.1007/7854_2016_459.

54.   Ffytche, D. H. Visual hallucinatory syndromes: past, present, and future. *Dialogues Clin Neurosci* **9**, 173–189 (2007).

55.   Netzer, O. *et al.* Trauma Under Psychedelics: MDMA Shows Protective Effects During the Peritraumatic Period. 2024.03.28.587237 Preprint at https://doi.org/10.1101/2024.03.28.587237 (2024).

56.   Chapman, L. J., Edell, W. S. & Chapman, J. P. Physical anhedonia, perceptual aberration, and psychosis proneness. *Schizophrenia Bulletin* **6**, 639 (1980).

57.   Hobson, J. A., Hong, C. C.-H. & Friston, K. J. Virtual reality and consciousness inference in dreaming. *Front. Psychol.* **5**, (2014).

58.   Windt, J. M. & Metzinger, T. The philosophy of dreaming and self-consciousness: What happens to the experiential subject during the dream state? in *The new science of dreaming: Volume 3. Cultural and theoretical perspectives* 193–247 (Praeger Publishers/Greenwood Publishing Group, Westport, CT, US, 2007).

59.   Johnson, M. W., Hendricks, P. S., Barrett, F. S. & Griffiths, R. R. Classic psychedelics: An integrative review of epidemiology, therapeutics, mystical experience, and brain network function. *Pharmacology & Therapeutics* **197**, 83–102 (2019).

60. Barrett, F. S., Johnson, M. W. & Griffiths, R. R. Validation of the revised Mystical Experience Questionnaire in experimental sessions with psilocybin. *Journal of Psychopharmacology* **29**, 1182–1190 (2015).

61. James, W. *The Varieties of Religious Experience: A Study in Human Nature*. (Routledge, 2003).

62. Blanke, O. *et al.* Neurological and robot-controlled induction of an apparition. *Current Biology* **24**, 2681–2686 (2014).

63. Arzy, S., Seeck, M., Ortigue, S., Spinelli, L. & Blanke, O. Induction of an illusory shadow person. *Nature* **443**, 287 (2006).

64. Blanke, O., Landis, T., Spinelli, L. & Seeck, M. Out-of-body experience and autoscopy of neurological origin. *Brain* **127**, 243 (2004).

65. Blakemore, S. & Frith, C. Self-awareness and action. *Current Opinion in Neurobiology* **13**, 219–224 (2003).

66. Sass, L., Pienkos, E., Nelson, B. & Medford, N. Anomalous self-experience in depersonalization and schizophrenia: a comparative investigation. *Consciousness and cognition* **22**, 430–441 (2013).

67. MacLean, K. A., Leoutsakos, J. S., Johnson, M. W. & Griffiths, R. R. Factor analysis of the mystical experience questionnaire: A study of experiences occasioned by the hallucinogen psilocybin. *Journal for the scientific study of religion* **51**, 721–737 (2012).

68. Griffiths, R. R., Richards, W. A., McCann, U. & Jesse, R. Psilocybin can occasion mystical-type experiences having substantial and sustained personal meaning and spiritual significance. *Psychopharmacology* **187**, 268–283 (2006).

69. Blom, J. D. Alice in Wonderland syndrome. *Neurol Clin Pract* **6**, 259–270 (2016).

70. Drori, G. & Salomon, R. Replicated behavioral and physiological evidence of explicit and implicit manipulation of the Sense of Reality (SoR). (2023) doi:10.17605/OSF.IO/BT3KD.

71. Schuetz, I. & Fiehler, K. Eye Tracking in Virtual Reality: Vive Pro Eye Spatial Accuracy, Precision, and Calibration Reliability. *J Eye Mov Res* **15**, 10.16910/jemr.15.3.3.

72. Bentall, R. The Illusion of Reality: A Review and Integration of Psychological Research on Hallucinations. *Psychological Bulletin* **107**, 82–95 (1990).

73. Schütt, H. H., Harmeling, S., Macke, J. H. & Wichmann, F. A. Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision Res* **122**, 105–123 (2016).

74. Feigin, H., Baror, S., Bar, M. & Zaidel, A. Perceptual decisions are biased toward relevant prior choices. *Sci Rep* **11**, 648 (2021).

75. Bell, V., Halligan, P. W. & Ellis, H. D. The Cardiff Anomalous Perceptions Scale (CAPS): a new validated measure of anomalous perceptual experience. *Schizophrenia bulletin* **32**, 366–377 (2005).

76. Loewy, R. L., Pearson, R., Vinogradov, S., Bearden, C. E. & Cannon, T. D. Psychosis risk screening with the Prodromal Questionnaire—brief version (PQ-B). *Schizophrenia research* **129**, 42–46 (2011).

77. Cicero, D. C., Neis, A. M., Klaunig, M. J. & Trask, C. L. The Inventory of Psychotic-Like Anomalous Self-Experiences (IPASE): Development and validation. *Psychological Assessment* **29**, 13–25 (2017).

78. Laeng, B., Sirois, S. & Gredebäck, G. Pupillometry: A Window to the Preconscious? *Perspect Psychol Sci* **7**, 18–27 (2012).

79. Urai, A. E., Braun, A. & Donner, T. H. Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nat Commun* **8**, 14637 (2017).

80.   van Kempen, J. *et al.* Behavioural and neural signatures of perceptual decision-making are modulated by pupil-linked arousal. *eLife* **8**, e42541 (2019).

81.   Kayser, C. & Shams, L. Multisensory Causal Inference in the Brain. *PLOS Biology* **13**, e1002075 (2015).

82.   Rohe, T. & Noppeney, U. Cortical Hierarchies Perform Bayesian Causal Inference in Multisensory Perception. *PLOS Biology* **13**, e1002073 (2015).

83.   Yinan Cao, Christopher Summerfield, Bruno Lucio Giordano, Hame Park, & Christoph Kayser. Causal Inference in the Multisensory Brain. *Neuron* **102**, 1076-1087.e8 (2019).

84.   Chancel, M., Ehrsson, H. H. & Ma, W. J. Uncertainty-based inference of a common cause for body ownership. *Elife* **11**, e77221 (2022).

85.   Ernst, M. O. & Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433 (2002).

86.   E, B., E, K., D, E., M, B. & L, T. van E. Seeing gray when feeling blue? Depression can be measured in the eye of the diseased. *Biological psychiatry* **68**, (2010).

87.   Kometer, M., Schmidt, A., Jäncke, L. & Vollenweider, F. X. Activation of serotonin 2A receptors underlies the psilocybin-induced effects on α oscillations, N170 visual-evoked potentials, and visual hallucinations. *J Neurosci* **33**, 10544–10551 (2013).

88.   Callaway, D. A., Perkins, D. O., Woods, S. W., Liu, L. & Addington, J. Movement abnormalities predict transitioning to psychosis in individuals at clinical high risk for psychosis. *Schizophrenia research* **159**, 263–266 (2014).

89.   Mittal, V. A. *et al.* Striatal volumes and dyskinetic movements in youth at high-risk for psychosis. *Schizophrenia Research* **123**, 68–70 (2010).

90.   Bernasconi, F. *et al.* Robot-induced hallucinations in Parkinson's disease depend on altered sensorimotor processing in fronto-temporal network. *Science Translational Medicine* **13**, (2021).

91.   Salomon, R. *et al.* Agency Deficits in a Human Genetic Model of Schizophrenia: Insights From 22q11DS Patients. *Schizophrenia bulletin* (2021).

92.   Krugwasser, A. R., Stern, Y., Faivre, N., Harel, E. V. & Salomon, R. Impaired sense of agency and associated confidence in psychosis. *Schizophrenia* **8**, 32 (2022).

93.   Gabitov, E., Lungu, O., Albouy, G. & Doyon, J. Movement errors during skilled motor performance engage distinct prediction error mechanisms. *Commun Biol* **3**, 1–16 (2020).

94.   Campbell, M. E., Sherwell, C. S., Cunnington, R., Brown, S. & Breakspear, M. Reaction time "mismatch costs" change with the likelihood of stimulus–response compatibility. *Psychonomic Bulletin & Review* **30**, 184–199 (2023).

95.   Salomon, R. *et al.* The insula mediates access to awareness of visual stimuli presented synchronously to the heartbeat. *Journal of Neuroscience* **36**, 5115–5127 (2016).

96.   Park, H.-D. *et al.* Transient modulations of neural responses to heartbeats covary with bodily self-consciousness. *Journal of Neuroscience* **36**, 8453–8460 (2016).

97.   Seth, A. K. & Tsakiris, M. Being a Beast Machine: The Somatic Basis of Selfhood. *Trends in Cognitive Sciences* **22**, 969–981 (2018).

98.   Rosas, F. E. *et al.* The entropic heart: Tracking the psychedelic state via heart rate dynamics. *bioRxiv* 2023–11 (2023).

99.   Skora, L. I., Livermore, J. J. A. & Roelofs, K. The functional role of cardiac activity in perception and action. *Neuroscience & Biobehavioral Reviews* **137**, 104655 (2022).

100. Hajcak, G., McDonald, N. & Simons, R. F. To err is autonomic: Error-related brain potentials, ANS activity, and post-error compensatory behavior. *Psychophysiology* **40**, 895–903 (2003).

101. Seth, A. K. Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences* **17**, 565–573 (2013).

102. Barrett, L. F. & Simmons, W. K. Interoceptive predictions in the brain. *Nature Reviews Neuroscience* (2015).

103. Aston-Jones, G. & Cohen, J. D. AN INTEGRATIVE THEORY OF LOCUS COERULEUS-NOREPINEPHRINE FUNCTION: Adaptive Gain and Optimal Performance. *Annu. Rev. Neurosci.* **28**, 403–450 (2005).

104. Gabay, S., Pertzov, Y. & Henik, A. Orienting of attention, pupil size, and the norepinephrine system. *Atten Percept Psychophys* **73**, 123–129 (2011).

105. Reimer, J. *et al.* Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nat Commun* **7**, 13289 (2016).

106. Lim, J. Z., Mountstephens, J. & Teo, J. Emotion recognition using eye-tracking: taxonomy, review and current challenges. *Sensors* **20**, 2384 (2020).

107. Zele, A. J. & Gamlin, P. D. The pupil: behavior, anatomy, physiology and clinical biomarkers. *Frontiers in Neurology* vol. 11 211 (2020).

108. Braem, S., Coenen, E., Bombeke, K., Van Bochove, M. E. & Notebaert, W. Open your eyes for prediction errors. *Cogn Affect Behav Neurosci* **15**, 374–380 (2015).

109. Van Der Wel, P. & Van Steenbergen, H. Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychon Bull Rev* **25**, 2005–2015 (2018).

110. Kok, P., Mostert, P. & de Lange, F. P. Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences* **114**, 10473–10478 (2017).

111. Harrison, W. J., Bays, P. M. & Rideaux, R. Neural tuning instantiates prior expectations in the human visual system. *Nat Commun* **14**, 5320 (2023).

112. Grossberg, S. How hallucinations may arise from brain mechanisms of learning, attention, and volition. *Journal of the International Neuropsychological Society* **6**, 583–592 (2000).

113. Wadsworth, B. Piaget's theory of cognitive development: An introduction for students of psychology and education. *McKay* (1971).

114. Piaget, J. & Cook, M. *The Child's Construction of Reality*. (Routledge & Kegan Paul Wellingborough, 1976).

115. Subbotsky, E. Magical thinking in judgments of causation: Can anomalous phenomena affect ontological causal beliefs in children and adults? *British Journal of Developmental Psychology* **22**, 123–152 (2004).

116. Bohil, C. J., Alicea, B. & Biocca, F. A. Virtual reality in neuroscience research and therapy. *Nature Reviews Neuroscience* **12**, 752–762 (2011).

117. Thurley, K. Naturalistic neuroscience and virtual reality. *Front. Syst. Neurosci.* **16**, (2022).

118. Aronov, D. & Tank, D. W. Engagement of neural circuits underlying 2D spatial navigation in a rodent virtual reality system. *Neuron* **84**, 442–456 (2014).

119. Ford, J. M., Roach, B. J., Faustman, W. O. & Mathalon, D. H. Synch Before You Speak: Auditory Hallucinations in Schizophrenia. *AJP* **164**, 458–466 (2007).

120. Stripeikyte, G. *et al.* Fronto-temporal disconnection within the presence hallucination network in psychotic patients with passivity experiences. *Schizophrenia bulletin* **47**, 1718–1728 (2021).

121.	Montagnese, M. *et al.* A Review of Multimodal Hallucinations: Categorization, Assessment, Theoretical Perspectives, and Clinical Recommendations. *Schizophr Bull* **47**, 237–248 (2020).