Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Immersion as inference: An active inference account of induced realities and its relation to the two-competitor model

Laura Renz
Saarland University & School of Engineering
Kyra L. GauthierDickey
University of Hawai‘i at M¯ anoa
Jonas Vibell
University of Hawai‘i at M¯ anoa
Alexander L. Francis
Purdue University
Farah I. Corona-Strauss
Saarland University & School of Engineering
Daniel J. Strauss
daniel.strauss@uni-saarland.de

Saarland University & School of Engineering

Research Article

Keywords: immersion, presence, active inference, free energy principle, precision, attention, virtual reality, augmented reality, neuroergonomics

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11647803/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Immersion as inference: An active inference account of induced realities and its relation to the two-competitor model

Laura Renz[1,2], Kyra L. GauthierDickey[3], Jonas Vibell[3], Alexander L. Francis[4], Farah I. Corona-Strauss[1,2], Daniel J. Strauss[1,2,*]

[1]Systems Neuroscience & Neurotechnology Unit, Faculty of Medicine,
Saarland University & School of Engineering, htw saar, Homburg/Saar, Germany

[2]Center for Digital Neurotechnologies Saar (CDNS),
Saarland University and htw saar, Germany

[3]Brain and Behavior Laboratory, Department of Psychology,
University of Hawai‘i at Mānoa, Honolulu, HI, United States

[4]Speech Perception & Cognitive Effort Lab, Department of Speech, Language,
and Hearing Sciences, Purdue University, West Lafayette, IN, United States

[*]Correspondence: daniel.strauss@uni-saarland.de

November 4, 2026

Abstract

Immersive technologies are designed for a single target: the human perceptual system. Engineers of virtual and augmented reality have long recognized that a display does not have to replicate the physical world, it only has to provide the inputs that satisfy the senses at the level at which the brain samples them. This engineering intuition is, in fact, a statement about brain function. Here we develop it into a formal account of immersion within the active inference framework. We treat an induced reality (a virtual, augmented, or mixed reality, but also a book or a daydream) as a candidate generative model that competes with the model of the actual reality for the explanation of a shared sensorium and for the control of action. Immersion is then defined as the ensuant belief that the induced reality best explains the current sensory signals and supports the current perceived reality of the winning model, despite any break in immersion by Bayesian model switching driven by a high-precision prediction error that the induced model cannot absorb. Exogenous and endogenous attention, implemented as weights governing stream selection in our earlier two-competitor model, which treated attention as necessary but not sufficient for immersion, appear as two routes of precision allocation: likelihood precision of evolutionarily entrenched sensory events, and precision that follows the expected free energy of policies. Affect and motivation enter through the pragmatic and epistemic value of remaining in the induced reality and through the rate of change of free energy. We show that the two-competitor attentional focus vector and its angle are recovered from the active inference variables, and we reproduce the two scenarios of the original model, a rewarding and a fatiguing activity interrupted by a slamming door, in a minimal simulation. The account links presence, embodiment, cybersickness, and narrative absorption to a single quantity, and it yields concrete predictions for the psychophysiological assessment and closed-loop control of immersion. We discuss implications for neuroergonomic design in which the evolutionarily entrenched core of perception is non-negotiable while its experience-dependent periphery can be trained.

Keywords: immersion; presence; active inference; free energy principle; precision; attention; virtual reality; augmented reality; neuroergonomics

1 Introduction

Interest in sensory-driven immersion has increased alongside advances in affordable head-mounted displays, spatial audio, haptic interfaces, and the business models built around a digital virtual universe [1]. Yet a broadly accepted definition of immersion in terms of established neurocognitive constructs and measurable quantities has remained elusive. In our earlier work, we approached this problem from the perspective of attention and proposed the two-competitor model [1]. The model treats an induced reality (a virtual, augmented, or mixed reality, but also a novel or a daydream) and the actual reality as two orthogonal dimensions competing for the projection of an attentional focus vector, and it integrates the taxonomy of exogenous versus endogenous attention with the taxonomy of external versus internal attention described by Chun et al. [2]. The angle of the attentional focus vector relative to the actual reality axis quantifies the degree of immersion, and its magnitude quantifies the total allocation of attentional resources. Affect and motivation enter the model through a fatigue module that maps reward and effort onto the endogenous weights [3, 4]. The central hypothesis of that work was that attention is a necessary but not a sufficient condition for immersion: without attention to the induced reality there is no immersion, but attention alone does not produce it, because the stimuli must also resonate with the goals and the affective state of the person. The two-competitor model captured this resonance only through the empirical weights of its fatigue module.

The two-competitor model was deliberately conceptual. It provided a taxonomy and a graphical intuition, and it allowed a quantitative implementation, but it left open many questions that a neuroscience of immersion must eventually answer: Why does an induced reality "win" the competition at all, given that the physical world never disappears from the sensorium? What does it mean, in terms of brain computation, for the stimuli of an induced reality to "fit together in some particularly compelling way" [1]? Why is immersion so fragile win respect to interruptions such as a a slamming door but so robust with respect to e.g. the weight of a headset? How do affect, motivation, and effort modulate immersion beyond the empirical weights that we assigned to them? Why is attention, which is plainly necessary for immersion, not sufficient for it? And how does immersion relate to the sensory conflict thought to contribute to cybersickness, which we have recently discussed as a signature of the evolutionary mismatch between our sensory systems and the technological environments of the Anthropocene [5, 6]?

In this article, we argue that all of these questions can be addressed within a common computational framework provided by active inference [7, 8]. The argument has a closely related engineering counterpart in immersive technology. Michael Abrash, who has led virtual reality research at Valve and later at Oculus and Meta Reality Labs, has consistently framed the goal of immersive technology not as replication of the physical world but as satisfaction of the perceptual system: "our conscious minds never actually interact with the real world. Instead, we interact with signals from sensors", humans are "inference machines, not objective observers", and "virtual reality doesn't have to replicate the real world in order to work. It just has to pro-

vide the right inputs [...] to satisfy the relevant senses" [9–11]. The technical requirements associated with presence that emerged from this view (field of view, display persistence, refresh rate, tracking precision, and motion-to-photon latency of about twenty milliseconds) were identified by investigating which perceptual subsystems must be satisfied before the brain accepts the displayed world as the place one is in, a process occurring "well below conscious awareness" [11]. This is an operational description of a brain that infers the causes of its sensations under a generative model, and that accepts the displayed world as the place one is in when the hypothesis that one is in that world explains the sensory input better than the hypothesis that one is wearing a display in a room. The free energy principle and active inference provide a framework for formalizing this interpretation [7, 8, 12], and predictive processing accounts of presence have already been proposed in the context of interoception [13, 14], sensorimotor contingencies [15, 16], phenomenal transparency [17, 18], body ownership [19, 20], and the brain as a simulator [21]. Most recently, it has been argued that the sense of presence in virtual reality and the sense of reality in everyday perception are mechanistically continuous [22], and electroencephalographic prediction error signatures have been used to detect implausible interactions in virtual reality [23–26].

What is missing, is a formal treatment that (a) represents the actual and the induced reality as competing hypotheses within one generative model, (b) derives the role of attention from precision, thereby explaining the attention taxonomies of the two-competitor model rather than assuming them, (c) integrates affect, motivation, and effort through expected free energy, (d) explains breaks in immersion, sensory conflict, and cybersickness within the same variables, and (e) connects to psychophysiological measurement and to the neuroergonomic design of immersive technologies. This is the goal of the present article. We first review the constructs of immersion and presence, followed by the necessary elements of the two-competitor model and of active inference (Section 2). We then develop the active inference account of induced realities (Section 3), map it onto the two-competitor model and reproduce its two scenarios in a minimal simulation (Section 4), and discuss implications for assessment, design, closed-loop systems, and neuroergonomics (Section 5) before addressing limitations and drawing conclusions.

2 Background

2.1 Immersion, presence, and induced realities

The literature on immersion is large and terminologically heterogeneous; see Agrawal et al. [27], Skarbez et al. [28], Nilsson et al. [29] for comprehensive reviews. Different theoretical approaches define immersion and its relationship to related concepts such as presence in different ways. Slater and colleagues reserve the term immersion for the objective properties of a system, the set of sensorimotor contingencies it supports, and use presence for the subjective response [30–32]. Within presence they distinguish place illusion, the illusion of being in

the place depicted, from plausibility, the illusion that the events depicted are really happening, and they emphasize that both illusions persist despite the sure knowledge that they are illusions [33, 34]. Similarly, Lombard and Ditton define presence as the perceptual illusion of nonmediation [35]. In contrast, Wirth and colleagues describe spatial presence as the outcome of a two-step process in which a spatial situation model is first constructed and then accepted as the primary egocentric reference frame [36, 37]. Finally, Agrawal and colleagues, whose definition we took as our starting point in the two-competitor model, define immersion as "a state of deep mental involvement in which the subject may experience disassociation from the awareness of the physical world due to a shift in their attentional state" [27, as quoted in 1]. Related constructs include embodiment, the sense of self-location, agency, and ownership over a virtual body [38–40], flow [41, 42], and narrative transportation [1, 43, 44]. What these definitions share is that the person is situated, by their own perceptual system, in one environment rather than another; where they differ is in what they take to be primary, the technology, the attentional state, the spatial model, or the illusion of nonmediation. Meta-analytic evidence indicates that technological immersion in Slater's sense increases presence with a medium effect size, with tracking, field of view, and stereoscopy among the most effective factors [45]. Assessment of presence and immersion has relied on physiological responses to stressful virtual environments [46], electroencephalographic correlates of spatial presence [47, 48], and counts of reported breaks in presence [49]. In this article we keep the two terms apart as follows. Immersion denotes the graded state of the person, the degree to which the induced reality, rather than the actual one, is taken to be the reality one is in, which is Agrawal's attentional definition made quantitative in Section 3.1. Presence denotes the subjective transparency of that state, the illusion of nonmediation of Lombard and Ditton and the place illusion of Slater, which Section 3.1 identifies with the absence of prediction error at the level of the context. Slater's immersion, the sensorimotor contingencies a system supports, is in our terms a property of the induced reality rather than of the person, and it enters the account as the evidence that the induced reality can supply. The predictive processing accounts reviewed in Section 2.4 concern presence in this sense, and they are the natural predecessors of an account in which immersion and presence are two aspects of one inference.

2.1.1 On actual, physical, and induced realities

In the two-competitor model we used the term induced reality to cover the whole range of situations in which attention is directed to a reality other than the one the person is actually in, from reading a book to a fully tracked multisensory virtual environment [1]. We continue to use the terms *induced reality* here, as well as *actual reality* rather than *physical reality* for its counterpart, because this distinction allows for cases in which physically generated cues contribute to the induced reality. A film screened with a live orchestra, a situation comedy with a laughter track or with the laughter of a physically present audience, and an extended reality installation with live-played music all integrate real-world cues into the induced reality, and

such cues may deepen immersion when they are congruent with it and disrupt it when they are not [50, 51]. Whether a cue is treated as part of the induced reality or as a competing signal is therefore decided by its coherence with the other elements of the scene, not by its physical origin, and once the cues are assigned, the attentional dynamics of the two-competitor model apply. This point is easily missed. In a recent position paper, Richardson, Gilchrist, and Stanton Fraser adopt a source-based interpretation and equate the actual reality with the physical world. They then conclude, from audience laughter, shared snacks, and printed books, that input from the actual reality (i.e., the physical world) can enhance immersion, which they take to contradict the orthogonal competition of the model [52]. Under this reading, the two-competitor model indeed fails. However, the term *actual reality* was chosen in the original article in preference to the more natural *physical reality* precisely to keep the two apart, but a loose, passing gloss of the actual reality as only and entirely the physical world around us and inside us [1] has invited the source-based reading. Under the model's own criterion, congruent laughter belongs to the comedy as much as the lines of the actors; the comfort of a printed book and the company of other viewers enter through the endogenous weights, where motivation was located from the start, and the study behind the laughter example shows the congruency dependence that this criterion predicts and a source-based reading cannot explain: laughter and screams raised reported immersion only when they matched the genre of the film [53]. Ultimately, an induced reality is in this sense always embedded in the actual reality: the display is a physical object, the reader sits in a room, and the brain continues to receive sensory information from the physical environment. Nevertheless, attention is directed toward one in preference to the other. The virtue of the terminology is that it forces a theory to explain what these situations have in common, and what distinguishes them, in a single set of variables. An active inference account meets this demand naturally, because it describes every one of these situations as inference under a generative model that receives sensory evidence of varying richness, and because it makes the assignment of a cue to a reality part of the inference itself (Section 3.5).

2.2 The two-competitor model

The two-competitor model represents the actual reality (ar) and the induced reality (ir) as orthogonal axes of a plane in which an attentional focus vector $\boldsymbol{a} = a_{\mathrm{ar}}\boldsymbol{e}_1 + a_{\mathrm{ir}}\boldsymbol{e}_2$ is mapped [1]. Its angle $\varphi_a = \tan^{-1}(a_{\mathrm{ir}}/a_{\mathrm{ar}})$ quantifies the degree of attention directed to the induced reality, with $\varphi_a = \pi/2$ indicating that attention is directed entirely to the induced reality, and its magnitude $m_a = \|\boldsymbol{a}\|_2$ quantifies the total attentional allocation. The projections are sums of exogenous and endogenous contributions,

$$a_{\mathrm{ar/ir}}(t) = \Omega_{\alpha,\beta,\gamma}\big(w^{\mathrm{exo}}_{\mathrm{ar/ir}}(t)\big) + \Omega_{\alpha',\beta',\gamma'}\big(w^{\mathrm{endo}}_{\mathrm{ar/ir}}(t)\big), \tag{1}$$

where $\Omega_{\alpha,\beta,\gamma}(x) = \alpha/(1+e^{-\beta x+\gamma})$ is a sigmoidal transfer function that maps exogenous or endogenous weights into a bounded allocation of attentional resources. The model is config-

ured to favor sufficiently strong exogenous signals, allowing abrupt and intense events to override ongoing immersion. The weights themselves were defined by two parameterized maps, $w^{\text{exo}} = \Phi_{\boldsymbol{p}}(\boldsymbol{s}(t))$ for the stimulus-driven properties of the streams (brightness, loudness, valence, arousal) and $w^{\text{endo}} = \Psi_{\boldsymbol{p}}(\boldsymbol{s}(t))$ for the goal-driven allocation that reflects internal attention, motivation, and affect, with a cognitive fatigue model [3, 4] as the worked example for $\Psi_{\boldsymbol{p}}$. Two scenarios illustrated the dynamics: a rewarding activity in the induced reality with a stable motivation profile, and an effortful, fatiguing activity in which motivation progressively declines, both interrupted after ten minutes by a slamming door in the actual reality that transiently pulls the attentional focus vector back toward the ar axis [1].

The model thus already contained four commitments that we now make explicit in active inference terms. First, attention is a necessary but not a sufficient condition for immersion. Second, immersion is a competition between two realities for a limited resource. Third, the competition is modulated by two modes of attentional allocation, exogenous (stimulus-driven) and endogenous (goal-directed) attention, both of which contribute to the allocation of attention between the two realities. Fourth, affect and motivation influence attentional allocation through the endogenous weights, which is where the insufficiency of attention was located. The maps $\Phi_{\boldsymbol{p}}$ and $\Psi_{\boldsymbol{p}}$ were placeholders for "any model which provides this" [1]. Active inference offers a framework meeting these requirements.

2.3 Active inference in brief

Active inference describes perception, attention, and action as consequences of a single imperative, the minimization of variational free energy under a generative model of how sensations are caused, with expected free energy guiding the selection of future policies [7, 8, 12, 54]. In more intuitive terms, the brain continuously predicts its sensory input and compares these predictions with incoming observations. Discrepancies can be reduced either by updating beliefs about the causes of sensory input (perception) or by acting to bring sensory input into closer agreement with predictions (action) [8]. For sensory observations o and hidden states s, the generative model $p(o,s)$ specifies the probabilistic relationships between hidden states and the observations they are expected to generate. Hidden states represent underlying conditions or causes, such as the source of a sound, which cannot be directly observed but must be inferred from sensory input. The exact posterior $p(s \mid o)$ describes the probability distribution over hidden states given the sensory observations and is here approximated by a distribution $q(s)$. Variational free energy measures how well beliefs about hidden states account for sensory observations. It provides an upper bound on surprise, the negative log-evidence of the observations under the model, and can be expressed as:

$$F[q,o] = \mathbb{E}_{q(s)}\big[\ln q(s) - \ln p(o,s)\big] = D_{\text{KL}}\big[q(s) \,\|\, p(s \mid o)\big] - \ln p(o) \;\geq\; -\ln p(o), \qquad (2)$$

Perception reduces F by updating beliefs about hidden states represented by $q(s)$, while

learning adjusts the parameters of the generative model $p(o,s)$. Action can change sensory observations by acting in the environment, thereby altering o. In hierarchical continuous-state formulations, the gradient of F includes terms involving precision-weighted prediction errors $\boldsymbol{\zeta}\boldsymbol{\varepsilon}$. Sensory prediction errors can be expressed as $\boldsymbol{\varepsilon} = o - g(s)$, where $g(s)$ represents the predicted sensory input, and describe discrepancies between predicted and observed sensory input. Precision $\boldsymbol{\zeta}$, the inverse of the expected variance of these errors, determines how strongly they influence inference [55–57]. Attention, in this framework, is precision optimization: increasing the precision of a sensory channel increases its influence on inference, and attentional effort may further be conceptualized in terms of the costs associated with maintaining precision estimates under uncertainty [6, 55, 58, 59].

The selection of future actions is guided by expected free energy, which evaluates possible courses of action in terms of their anticipated outcome and expected information. For a policy ξ, a sequence of actions, the expected free energy $G(\xi)$ of its future outcomes decomposes as

$$G(\xi) = \underbrace{-\mathbb{E}_{q(o_\tau|\xi)}\big[\ln p(o_\tau \mid C)\big]}_{\text{negative pragmatic value}} \underbrace{-\mathbb{E}_{q(o_\tau|\xi)}\Big[D_{\mathrm{KL}}\big[q(s_\tau \mid o_\tau,\xi)\,\|\,q(s_\tau \mid \xi)\big]\Big]}_{\text{negative epistemic value}}, \tag{3}$$

where C encodes prior preferences over outcomes [8, 12, 60]. In the absence of additional policy priors, policies are selected according to $q(\xi) = \sigma(-\gamma_\xi\, G(\xi))$, a softmax over policies with precision γ_ξ; for two policies the softmax reduces to the logistic function $\sigma(x) = 1/(1+e^{-x})$ of the difference in expected free energy, and we use σ for both in what follows. The first term captures pragmatic value, reflecting how well the anticipated outcomes of a policy align with prior preferences. The second captures epistemic value, reflecting the expected information gain, or reduction of uncertainty about hidden states [58, 60, 61]. Because G is the sum of the negative pragmatic and the negative epistemic value, higher pragmatic or epistemic value leads to lower expected free energy. The epistemic term thereby formalizes intrinsic motivation, curiosity, and exploration, and it has been used to explain why agents seek novelty and why the epistemic value of an activity collapses once it no longer resolves uncertainty [61, 62].

Two further elements are needed. First, generative models are hierarchical and deep in time: slow, abstract states at higher levels contextualize fast, concrete states at lower levels, so that a narrative or a goal can constrain the interpretation of sensory input over seconds to hours [63–65]. In our earlier work we represented this hierarchy as a spatiotemporal scale space with an evolutionarily entrenched, fast, and near core and a slower, reasoning-based periphery [5, 6], and we distinguished within the core a species-typical, genetically entrenched set of biases from a plastic, experience-dependent mantle [5]. Second, affect has been given a free energy account. Joffily and Coricelli proposed that valence tracks the rate of change of free energy, negative when free energy rises and positive when it falls [66]; Hesp and colleagues modeled valence as an inference based on the expected precision of one's own action model [67]; and Kiverstein and colleagues described the feeling of grip as the experience of prediction errors

being reduced at a rate that is better than expected [62, 68]. Interoceptive inference extends the framework to the body: emotions and the sense of a bodily self arise from predictions about the causes of interoceptive signals, and the precision of these predictions determines their contribution to experience [14, 69–72].

Finally, and central to what follows, Parvizi-Wayne has argued that the dichotomy between exogenous and endogenous attention becomes less clear within active inference: preferences "enslave" attention, so that a stimulus-driven capture and a goal-driven allocation are both instances of precision being assigned to the streams that matter for the expected free energy of the agent's policies, the two differing, in our reading of his argument, in the level and time scale of the model at which the precision is set [73]. We used this argument for effortful listening in the soundscapes of the Anthropocene [6] and for sensory conflict and strain in neuroergonomics [5]. Here we use it to reinterpret the exogenous and endogenous contributions to attentional allocation in the two-competitor model.

2.4 Predictive processing accounts of presence

Predictive processing accounts have so far addressed presence, in the sense fixed in Section 2.1, rather than immersion. Seth, Suzuki, and Critchley proposed an early predictive processing model of presence, in which the sense of presence arises when interoceptive prediction errors are successfully suppressed by top-down interoceptive predictions. This suggests that presence depends in part on whether bodily responses can be integrated into a coherent prediction of the current situation [13]. The model was later tested in virtual reality, where only high-arousal emotional states modulated presence [74]. Building on the model, Deane et al. discussed how the sense of presence is diminished in depersonalization [75]. Seth also proposed that perceptual presence, the phenomenal sense that an object is really there, depends on the counterfactual richness of the generative model, that is, on how many sensorimotor contingencies the model encodes for the object [15, 76], an idea rooted in the sensorimotor account of perception [16] and directly relevant to Slater's characterization of immersion as the set of valid sensorimotor contingencies a system supports [31]. Building on Metzinger's account of phenomenal transparency, the fact that we normally experience the world rather than our model of it [77], Limanowski and Friston related this phenomenon to the precision assigned to the model itself: a model that is used for inference with high confidence is transparent, whereas a model whose precision is lowered becomes opaque, an object of attention in its own right [17, 78]. Body ownership illusions have been modeled as Bayesian causal inference over the common cause of visual, tactile, and proprioceptive signals [19, 20, 39]. Riva and colleagues described the brain as a simulator whose predictions, when matched by the technology's predictions about the sensory consequences of the user's actions, produce presence [21, 79], and Gonzalez-Franco and Lanier proposed a model of illusions in virtual reality based on bottom-up and top-down processing [80]. Recent electroencephalographic work has identified prediction error signatures

of implausible visuo-haptic interactions in virtual reality and their modulation by presence and embodiment [23, 25, 26], complementing our own observation of a cross-modal semantic incongruency response to mismatching tactile and virtual visual experiences [24]. Singleton and Fotopoulou have recently argued that presence in virtual reality and the sense of reality in everyday perception share a predictive processing mechanism [22], and reality monitoring has been described as a comparison of subjective signal strength to a reality threshold [81, 82].

Taken together, these accounts suggest that presence can be understood as emerging from perceptual and inferential processes rather than as a separate component added to perception. To our knowledge, none of these accounts combines an explicit competition between induced and actual reality with a precision-based treatment of both attention taxonomies and an account of the affective and motivational dynamics that govern immersion. The two-competitor model contains exactly these elements in descriptive form. The following section brings the two together.

3 An active inference model of induced realities

3.1 The reality as a hidden context

We begin with the observation that the sensorium of an immersed person is never generated by a single reality. A person wearing a head-mounted display may receive visual and auditory streams from the induced reality, but also the weight of the headset, the sound of the room, the pressure of the chair, hunger, fatigue, and, crucially, proprioceptive and vestibular signals that are generated by the physical body in the physical room. A reader receives the words of the novel and, at the same time, the light of the lamp and the noise of the street. The question for the brain is therefore not which reality is present, but which reality provides the best explanation of the sensory signals it is currently sampling and the best basis for the policies it is currently pursuing.

We formalize this by introducing a hidden context variable $c \in \{\mathrm{ar}, \mathrm{ir}\}$ as one of the slow, abstract states that, in the hierarchy of Section 2.3, contextualize the faster states and the observations below them,

$$p(\boldsymbol{o}_{1:T}, \boldsymbol{s}_{1:T}, c, \xi) = p(c)\, p(\xi)\, p(\boldsymbol{s}_1 \mid c)\, p(\boldsymbol{o}_1 \mid \boldsymbol{s}_1, c) \prod_{t=2}^{T} p(\boldsymbol{o}_t \mid \boldsymbol{s}_t, c)\, p(\boldsymbol{s}_t \mid \boldsymbol{s}_{t-1}, \xi, c), \qquad (4)$$

in which $\boldsymbol{o}_{1:T} = (\boldsymbol{o}_1, \ldots, \boldsymbol{o}_T)$ and $\boldsymbol{s}_{1:T}$ denote the sequences of observations and hidden states over the T time steps of an episode, and in which the context selects the likelihood mapping $p(\boldsymbol{o}_t \mid \boldsymbol{s}_t, c)$ and the transition dynamics $p(\boldsymbol{s}_t \mid \boldsymbol{s}_{t-1}, \xi, c)$, that is, the sensorimotor contingencies and the narrative and causal structure that the agent uses to predict its sensations. If the induced reality provides the relevant context ($c = \mathrm{ir}$), predictions are based on the induced reality's model: head rotation predicts a rotation of the displayed scene, a virtual approach predicts a

change in binaural cues, and a plot predicts the next event. Under $c = \text{ar}$, the model of the actual reality is engaged: head rotation predicts corresponding shifts of the room one is in, and the plot is irrelevant. Both models make predictions about the same observation vector $\boldsymbol{o}_t$, which we partition into modality-specific streams o_t^m, $m = 1, \ldots, M$, including exteroceptive streams delivered by the induced reality, exteroceptive streams that only the actual reality generates, and bodily streams (proprioception, vestibular signals, interoception) that are generated by the physical body under either context. Figure 1 summarizes the structure.

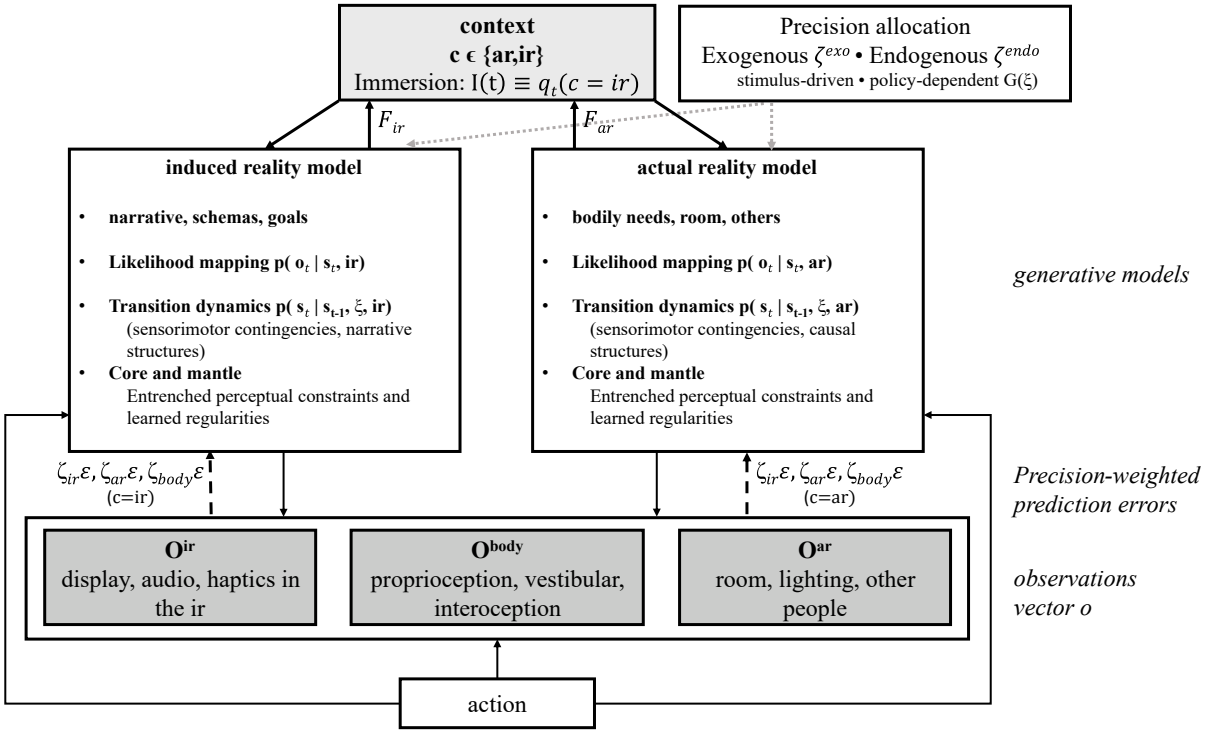


Figure 1: Schematic of the generative model of induced realities. A context variable c selects which of two models, that of the induced reality (ir) or that of the actual reality (ar), is used to predict a shared observation vector consisting of ir-specific streams, ar-specific streams, and bodily streams that are generated by the physical body under either context. Precision-weighted prediction errors contribute to the variational free energy of each context, F_{ir} and F_{ar}, which informs context inference and the degree of immersion $I(t) = q_t(c = \text{ir})$. Solid arrows denote predictions, dashed arrows precision-weighted prediction errors and dotted arrows precision control.

The two context-dependent models are not entirely independent, but share parts of their generative structure, with context-specific differences in the likelihood mappings and transition dynamics. One implication of this shared structure is that an induced world need not be constructed from scratch, an idea closely related to Ryan's principle of minimal departure. According to this principle, fictional worlds are assumed to resemble the actual world as closely as possible, departing from it only where the narrative requires. A reader encountering a fictional world in which sorcerers exist will still assume that gravity operates as usual, even if the author never mentions it [83]. This helps explain why a novel can evoke a rich imagined world from relatively little sensory input by drawing on existing knowledge. This shared generative structure also provides a possible basis for interactions between the two realities: prior experi-

ence shapes the interpretation of induced worlds, while learning within an induced reality may update knowledge or skills that subsequently transfer to the actual reality. For example, information encountered in fiction can influence real-world beliefs [43, 84, 85], and skills acquired through VR training can transfer to real-world tasks [86, 87]. Such cross-context learning is not explicitly modeled here.

Inference on the context follows from the free energy, Eq. (2), of the two sub-models. Writing F_c for the variational free energy obtained when the observations are explained under context c, the posterior over the context is

$$q(c) \propto p(c)\exp(-F_c), \qquad \lambda \equiv \ln\frac{q(c=\mathrm{ir})}{q(c=\mathrm{ar})} = \ln\frac{p(\mathrm{ir})}{p(\mathrm{ar})} + F_{\mathrm{ar}} - F_{\mathrm{ir}}, \tag{5}$$

so that the log-odds λ of being in the induced reality are given by the difference in free energy, or equivalently in log-evidence, between the two explanations. Equation (5) is an instantaneous posterior. If the context is treated as a hidden state that changes slowly, with a transition prior $p(c_t \mid c_{t-1})$ that favors staying, the prior term becomes the prediction from the previous posterior and the evolution of λ can be approximated by a recursive, leaky accumulation of the free energy difference over time, which is the form we use in the simulation of Section 4.2. Because free energy is a sum of precision-weighted prediction errors over streams and hierarchical levels, the difference $F_{\mathrm{ar}} - F_{\mathrm{ir}}$ is dominated by the streams to which precision is assigned. We use λ, and its logistic transform $q(c=\mathrm{ir}) = \sigma(\lambda)$, as the central quantity of this article.

Definition of immersion. We define the degree of immersion at time t as the posterior probability that the induced reality is the context that best explains the current sensorium and that supports the current policy, $I(t) \equiv q_t(c=\mathrm{ir})$. This definition satisfies the requirements we set out for the two-competitor model. It is continuous and dynamic; it does not require introspection, since $q(c)$ is a state of the inference machinery rather than a belief one can report on without changing it; and it covers the whole range of induced realities, because the evidence that enters Eq. (5) can come from rich exteroceptive streams, from a narrative maintained at deep levels of the model, or from both. Moreover, the framework allows the endogenous influence of affect and intrinsic motivation, central to the original two-competitor model, to be represented through preferences, expected information gain, and their effects on policy selection and attentional precision.

Definition of presence. Following Limanowski and Friston [17] and Metzinger [77], we define presence as the transparency of the winning model: the induced reality's model is used to generate predictions and to guide action with high precision, while the context itself does not become an object of attention. In the terms of Eq. (5), presence corresponds to a regime in which λ is large and, importantly, stable, so that the induced model is used with high confidence and the context itself generates no prediction errors that would attract attention. When λ

approaches zero, the context becomes uncertain, prediction errors at the context level grow, precision is drawn to the question "where am I?", the mediation becomes opaque, and one notices the headset, the page, or the screen. This distinguishes presence from immersion: presence is the absence of prediction error at the context level, immersion is the outcome of the context inference. Abrash's remark that presence "is distinct from immersion, which merely means that you feel surrounded by the image of the virtual world; presence means that you feel like you're in the virtual world" [11] is consistent with this distinction.

Place illusion and plausibility. Slater's two orthogonal illusions primarily map onto two components of F_{ir} [31, 33]. Place illusion is the success of the induced model at the level of sensorimotor contingencies, that is, the smallness of prediction errors in $p(\boldsymbol{o} \mid \boldsymbol{s}, \mathrm{ir})$ when the agent acts: turning the head, reaching, walking. Plausibility is the success of the induced model at the level of scene and narrative states, that is, the smallness of prediction errors in $p(\boldsymbol{s}_t \mid \boldsymbol{s}_{t-1}, \xi, \mathrm{ir})$: events unfold as the model of a coherent world predicts, and the world responds to the agent as a world would. Both contribute to the free energy difference in Eq. (5), and both can be lost independently, as Slater and colleagues have documented.

3.2 Precision and the two taxonomies of attention

In the two-competitor model, the exogenous and endogenous weights were given by the maps $\Phi_{\boldsymbol{p}}$ and $\Psi_{\boldsymbol{p}}$. In active inference both weights are precisions, and the maps are replaced by the mechanisms that set precision. Following Parvizi-Wayne [73], we distinguish two routes of precision control, which we write as ζ_m^{exo} and ζ_m^{endo} for stream m, not because they are different kinds of attention but because they are set at different levels of the model and on different time scales.

Exogenous precision. The likelihood precision of certain sensory events is high by construction of the model, and the construction is evolutionary. Abrupt onsets, looming, sudden loudness, motion in the periphery, faces, and the interruption of a regularity carry high precision because, in the environments in which the human perceptual system evolved, they were the events whose neglect was most costly [5, 6, 88–90]. These precisions are part of the genetically entrenched core of the model, they are fast, they are modality-specific, and they are largely independent of the current context. In the two-competitor model we captured this with a transfer function that gives priority to the exogenous channel, and in the fatigue scenario we represented the slamming door as a brief exogenous spike. Here the door is a high-precision prediction error in an ar-specific stream, and it enters $F_{\mathrm{ar}} - F_{\mathrm{ir}}$ with a weight that the induced model cannot reduce because the induced model does not predict doors in the physical room. The sound is better explained by the actual than by the induced model. Note that the induced reality has its own exogenous events (a jump scare, a virtual explosion), and these carry the same entrenched

precision, which is why they are so effective at pulling and holding attention [1].

Endogenous precision. The second route sets precision according to the expected free energy of policies. An agent that has selected the policy "continue reading the novel" or "continue the game" assigns high precision to the streams whose prediction errors are informative for that policy and low precision to the others. This is what Parvizi-Wayne means by preferences enslaving attention [73], it is what Feldman and Friston [55] modeled as attention, and it is the active inference reading of the biased competition and probabilistic stream selection models that we used as the basis of the two-competitor model [91–93]. Formally, we let the endogenous precision of the streams of reality c follow the posterior probability of the policies that act in c. Rather than determining endogenous precision independently for each stream, we first define a common precision at the level of each reality c. This context-level precision ζ_c^{endo} is determined by the posterior probability of policies in c and is assigned to the streams associated with that context. We adopt this simplification in the simulation presented, see Appendix A.

$$\zeta_c^{\text{endo}}(t) = \zeta_0 + (\zeta_{\max} - \zeta_0)\, q_t(\xi \in \Xi_c), \qquad q_t(\xi) = \sigma\big(-\gamma_\xi\, G_t(\xi)\big), \tag{6}$$

External and internal attention. The second taxonomy of the two-competitor model, external versus internal attention [2], does not require a third route.

Instead, it distinguishes the target of attentional selection. In external attention, precision is assigned to currently available sensory information, while in internal attention, precision is directed toward internally generated or maintained representations, such as memories, imagery, narratives, and simulations. The distinction cuts across the hierarchy of the generative model and can concern both lower sensory and higher interpretive levels. In the competition between actual and induced realities, both forms of attention can contribute to either reality. A novel, for example, engages external attention through the perception of written words and internal attention through the imagined narrative world. The relative contribution of external and internal attention to the induced reality therefore depends on the precision assigned to the respective representations.

How this precision is distributed may also depend on the sensory modality involved. Within the two-competitor model, internally generated and externally driven signals may compete for attentional precision within the same sensory modality. Vivid visual imagery could therefore reduce the attentional gain assigned to external visual input. Conversely, a sufficiently salient external stimulus might shift this balance toward the actual reality, potentially disrupting immersion more strongly when it engages the same modality as the induced representation than when it engages a different modality. The contribution of internally generated imagery may also vary between individuals, consistent with findings of reduced narrative absorption and engagement in individuals with aphantasia [94, 95].

Consequences. Three consequences follow. First, exogenous and endogenous precision control are not independent: exogenous precision is set by the model's construction, endogenous precision by the model's current policies, and the second can modulate the effect of the first only within limits set by the core. This is why one can learn to ignore the hum of an air conditioner but not a slamming door. It is the phenomenon that we have studied as long-term or attentional habituation, the drift of attention away from a repeated stimulus that is neither informative nor behaviorally relevant [96–98]. In the present terms such a stimulus carries little exogenous weight, and once its regularity has been learned it produces no prediction error and no information gain, so that the precision assigned to its stream fades and the stream drops out of the competition. In our habituation paradigm the phase locking of single-trial late auditory evoked responses declined over the repetitions of a soft tone but remained stable over the repetitions of a loud one [96–98], and we modeled this as a hippocampal comparator that detects novelty and withdraws attentional binding from familiar input [97]. A loud or aversive stimulus keeps the exogenous precision of the core and resists habituation, and a change in stimulus properties restores precision at once, which is dishabituation. Second, the competition between realities is decided by precision-weighted evidence, not by evidence alone. A headset can deliver a perfectly coherent scene and still lose the competition if the policies of the agent lie in the actual reality, and a novel can win the competition despite providing little exteroceptive evidence for the depicted world if the policies lie in the narrative. Third, the total precision that an agent can maintain is limited, and its maintenance costs effort [3, 6, 99, 100]. The magnitude m_a of the two-competitor model, the total attentional allocation, corresponds to the total precision currently deployed across both realities. The effort associated with maintaining and adjusting this precision allocation relates to what we have called sensory strain in the neuroergonomic context [5].

3.3 Affect, motivation, and the expected free energy of staying

The two-competitor model introduced affect through a fatigue module that mapped reward and effort onto the endogenous weight of the induced reality [1, 4]. Active inference offers a more general and more principled account. The endogenous precision of the induced reality follows, through Eq. (6), the expected free energy of the policies that act in it, and Eq. (3) decomposes this expected free energy into a pragmatic and an epistemic component. Both have direct readings in the phenomenology of immersion.

The pragmatic value of staying in the induced reality is the expected fulfillment of prior preferences: the reward of the game, the pleasure of the story, the social contact in a virtual meeting, the completion of a training task. The pragmatic value of leaving is the expected fulfillment of the preferences of the actual reality: eating, resting, answering the person at the door, the "increasing pressure to do something else in the actual reality" of our earlier scenario [1]. As bodily needs change during prolonged immersion, policies directed toward the actual

reality may gain pragmatic value. Sustaining immersion then depends on whether the combined pragmatic and epistemic value of remaining in the induced reality continues to outweigh that of competing policies.

The epistemic value of staying is the expected information gain, that is, the reduction in uncertainty about hidden states that the agent anticipates from remaining in the induced reality. A well-designed game, a well-told story, and a well-structured training scenario keep this term high by continuously posing questions that the agent can answer through action and observation. The two-competitor scenario of a game that becomes too hard is, in these terms, a joint collapse of pragmatic and epistemic value: the reward-to-effort ratio deteriorates, and predictions are made and violated but the violations cannot be resolved by any available policy, so that expected information gain, and with it the endogenous precision of the induced reality, declines. This is a different failure from boredom, in which epistemic value collapses because nothing is left to resolve, but both tend to reduce immersion [61, 62]. The region between the two failures, in which prediction errors are continuously generated and continuously resolved at a rate that matches the agent's capacities, is the region of flow [41, 42, 62, 68], and it is the region in which the endogenous precision of the induced reality is maximal.

Valence itself we read, following Joffily and Coricelli [66], as a function of the rate of change of free energy, positive when variational free energy falls and negative when it rises, in the spirit of the accounts that relate valence to error dynamics that are better or worse than expected [62, 67]. This gives a natural account of the affective texture of immersion. A rewarding activity in the induced reality produces a positive valence while its pragmatic value rises and prediction errors are resolved, and a neutral valence once a stable level is reached; a fatiguing activity produces a sustained negative valence as pragmatic and epistemic value decline; and a break in immersion produces a sharp negative transient, the startle, when an unexplained high-precision prediction error appears, followed by a positive transient as that error is resolved. Arousal, in turn, is the interoceptive correlate of a change in precision, which is consistent with the finding that presence is modulated by high-arousal emotional states [13, 74].

3.4 Breaks in immersion as Bayesian model switching

Equation (5) makes a break in immersion a Bayesian model comparison. As long as the induced model explains the precision-weighted sensorium at least as well as the actual model, λ remains positive and the agent stays immersed. A break occurs when a prediction error arises that the induced model cannot absorb and that carries enough precision to reverse the sign of $F_{\mathrm{ar}} - F_{\mathrm{ir}}$. Three kinds of event have this property. Exogenous events in the actual reality (the slamming door) carry entrenched high precision. Violations of the induced reality's own sensorimotor contingencies (latency, tracking loss, a hand that passes through a table, a mismatch between tactile and visual experience) generate prediction errors in streams to which the agent has assigned high endogenous precision, precisely because it is acting in the induced reality

[23, 24, 26, 101]. And violations of plausibility at the narrative level (a character who behaves inconsistently, an event without cause) generate prediction errors at the deep levels of the model.

The dynamics of a break follow from the precision-weighted update of λ. Because the rate of belief updating scales with the precision of the evidence [12, 55], a high-precision event switches the context quickly, on the time scale of an orienting response, whereas the re-accumulation of evidence for the induced reality after the event is slower, on the time scale of the endogenous precision. This asymmetry is what our earlier simulation produced with an oscillatory spike and a sigmoidal transfer function; here it is partly a property of the precision-dependent update rate and partly a property of the events themselves, since the exogenous transient is brief whereas the uncertainty it raises (what was that, and does it require action?) decays more slowly. The accumulation of evidence does not by itself make a long-immersed agent harder to break out than a newcomer; such dwell-time effects arise, in the present account, from learning, because time spent in the induced reality lets the agent learn its contingencies, which lowers F_{ir} and raises the disturbance required for a switch. We do not simulate learning here. Slater and Steed's virtual presence counter, which models presence as transitions between two states driven by breaks [49], is a discrete version of this picture.

Breaks are accompanied by orienting. In the auditory modality, we have shown that the vestigial auriculomotor system, the periauricular muscles that once oriented the pinna, is activated by the direction of exogenous and endogenous auditory attention [102, 103] and by effortful listening [104], so that a non-invasive electromyogram of these muscles provides a marker of the reflexive orienting that accompanies an exogenous override. The auriculomotor system also responds to affective events [105, 106]; we return to both aspects in Section 5.1.

3.5 Sensorimotor coherence and causal inference

In the two-competitor model we proposed that a set of stimuli supports immersion when they "fit together in some particularly compelling way" and that this coherence is what internal attention maintains [1]. Active inference gives this a precise meaning: the stimuli fit together when they are the outputs of a single generative model in which action has predictable consequences. The induced reality's model is engaged by the agent's own actions. Every head movement is a test of the hypothesis $c = \mathrm{ir}$, and the hypothesis survives only if the displayed scene responds as a world would. This is why tracking and latency dominate the technical requirements for presence [10, 11, 101, 107], why interactivity increases immersion [1, 45], and why Slater's definition of immersion as valid sensorimotor contingencies is the right definition of what a system must provide [31, 34]. Seth's counterfactual richness [15] adds that the sense of reality of an object increases with the number of sensorimotor contingencies the model encodes for it: an object one can walk around and pick up is more present than one that can only be looked at.

The same logic settles how a stream is assigned to a reality (Section 2). Every stream enters both F_{ir} and F_{ar} in Eq. (5), and it counts as evidence for whichever model predicts it

under the current precisions, irrespective of its physical origin. The live orchestra in the concert hall is predicted by the induced model of the film and adds evidence for $c = \mathrm{ir}$; a telephone ringing in the same hall is not, and adds evidence for $c = \mathrm{ar}$. Congruency, in this reading, is predictability under the induced model, and the assignment is a causal inference over the source of a signal of the same kind that decides whether multisensory signals share a common cause before they are integrated [108, 109]. Multisensory integration and immersion are therefore the same computation at two levels of the hierarchy: within a reality, the brain decides whether two cues share a cause and, if they do, fuses them weighted by their precisions [108, 110]; across realities, Eq. (5) decides whether the streams share a reality. The factors that break binding at the lower level, temporal asynchrony beyond the binding window, spatial offset, and semantic incongruency, are for this reason the factors that break immersion at the upper level, which answers the question we left open in the multimodal section of the two-competitor model [1]. In this account, the assignment of sensory streams to the two contexts is itself inferred rather than given by the stimuli, and the axes of the two-competitor plane are hypotheses about the causes of the sensorium.

The bodily streams occupy a special place. Proprioception, vestibular signals, and interoception are generated by the physical body under either context, so that the induced model must explain them or must at least not be contradicted by them. A seated user whose avatar sits is consistent; a seated user whose avatar flies generates a vestibular prediction error under the induced model that cannot be resolved by action. This is the sensory conflict of cybersickness [111–113], which we have argued arises from a deeply entrenched core of the model that predicts the covariation of visual and vestibular signals and that cannot be retrained on the time scale of a session [5, 6]. Within the present model, cybersickness is free energy that neither perception nor action can reduce: the induced model predicts the visual stream but not the vestibular stream, the actual model predicts the vestibular stream but not the visual stream, and no policy in either reality resolves the conflict. Formal treatments of visuo-proprioceptive conflict [114] and of sensory attenuation and illusions [115] in active inference, and of the difference between virtual and physical head pose [116] are consistent with this reading, as is the empirical finding that presence and cybersickness are negatively related [113]: the same prediction error that causes sickness reduces λ.

Embodiment follows the same logic. Ownership of a virtual body arises when the common-cause hypothesis, that the seen body and the felt body are one, has more evidence than the separate-cause hypothesis [19, 39, 40, 117], and full-body illusions in virtual reality show that this inference tolerates differences in the size, age, and appearance of the body as long as a first-person perspective, spatial coincidence, and sensorimotor and multisensory contingencies are preserved [118–122]. An embodied avatar increases immersion because it moves the bodily streams from the “not contradicted” category into the “explained” category of the induced model, and it increases sensitivity to violations of sensorimotor contingencies for the same reason [26].

3.6 Interoception, the core, and the mantle

Two further elements of our earlier work map onto the model. First, the interoceptive predictive coding model of presence [13] concerns the bodily streams in Figure 1: presence is supported when the induced model successfully predicts, or at least does not contradict, the current interoceptive state, so that arousal is attributed to the induced world. A horror scene that raises the heart rate is consistent with the induced model, which predicts fear; a horror scene watched while worrying about an email is not, because the arousal is attributed to the actual reality.

Second, the distinction between the core and the mantle of the perceptual system [5] provides a way of considering how readily different perceptual processes can be modified through experience. The core refers to the evolutionarily shaped, species-typical architecture that constrains how sensory information is processed and weighted and, consequently, how perception and attention are directed. It includes, for example, the omnidirectional monitoring of the auditory system, vestibular–proprioceptive coupling, the priority of foveal vision, the multisensory mapping of peripersonal space, and rapid cross-modal orienting to salient events. The mantle, in contrast, comprises those aspects of perception that are more readily shaped by experience, such as learned contingencies of a particular interface, tolerance for a certain latency, or habituation to a recurring background sound [5]. The temporal window within which multisensory signals are bound is a mantle property as well. The brain maintains the perception of synchrony across modalities whose transduction latencies differ, and it expects a larger gap between the visual and the auditory component of a distant event than of a near one, as between lightning and thunder [123]; the window also narrows with training [124]. The coupling of visual and vestibular signals, by contrast, belongs to the core. For a theory of immersion, this distinction is relevant in both the induced and the actual reality, because the extent to which a stimulus can be ignored, accommodated, or given attentional priority determines whether it supports the induced reality or competes with it. Some components of attention show substantial biological constraints and inter-individual genetic variation; for example, genetic variants have been associated with differences in visual attention [125], and genetic influences have also been reported for the emotion-modulated startle response [126]. This does not imply that such processes are unchangeable, but rather that not all perceptual and attentional tendencies are equally susceptible to adaptation. For immersive design, this means that not every discrepancy should be expected to become equally tolerable with experience. A VR game can assume that players will learn an initially unfamiliar mapping between controller input and virtual action, but it should not assume that repeated, unexpected sounds from the actual environment will become equally easy to ignore. Conversely, sudden onsets can deliberately be used within the induced reality to direct attention toward relevant events, while comparable events originating from the actual environment may compete with the induced reality and potentially disrupt immersion. An immersion system should therefore accommodate relatively constrained perceptual tendencies while exploiting the adaptability of more experience-dependent processes. This asymmetry

forms the basis of the design principle discussed in Section 5.2.

4 Reading the two-competitor model through active inference

4.1 Correspondence

Table 1 lists the elements of the two-competitor model and their counterparts in the present account. The correspondence is not one of replacement: the two-competitor model remains the appropriate level of description for experimental design and for the graphical intuition, and its quantities can be computed from the active inference variables. The endogenous weights $w^{\text{endo}}_{\text{ar/ir}}$ are the endogenous precisions of Eq. (6), set by expected free energy; the exogenous weights $w^{\text{exo}}_{\text{ar/ir}}$ are the entrenched likelihood precisions of the events currently present in each reality; the transfer functions Ω implement a sigmoidal saturation of precision into a bounded resource allocation; the projections $a_{\text{ar/ir}}$ and hence φ_a and m_a follow from Eq. (1). What the active inference account adds is the quantity $q(c = \text{ir})$, the perceptual outcome of the competition, which is related to the attentional focus vector but not identical to it: the angle φ_a describes where precision is currently deployed, $q(c = \text{ir})$ describes which reality has won the explanation. The two can dissociate, for instance immediately after a break, when precision has returned to the induced reality but the context has not yet been restored.

4.2 An illustrative simulation

To illustrate how the active inference account can reproduce the behavior of the two-competitor model and to make the dependencies explicit, we implemented a simplified version of the model and repeated the two scenarios of Strauss et al. [1]: a rewarding activity in the induced reality and an effortful, fatiguing activity, both interrupted after ten minutes by a slamming door in the actual reality. The implementation follows Eqs. (3) to (6) with the following simplifications. Rather than computing pragmatic and epistemic values from a full generative model, we specified how these values change over time for the policies associated with each reality: in scenario A the pragmatic value of the induced reality rises to a stable level and its epistemic value remains constant, whereas in scenario B both decline as the game becomes too hard, and in both scenarios the pragmatic value of the actual reality rises slowly with bodily needs. Endogenous precisions followed the policy posterior according to Eq. (6). Exogenous precisions were modeled, as in the original, by oscillatory functions, with an up-chirp in the induced reality of scenario B to represent the increasingly rapid stimulation of the difficult game, and the door was represented as a transient increase in the exogenous precision of an ar-specific stream together with a transient epistemic value of the actual reality (the question “what was that?”). Rather than computing the context posterior from a full generative model at each time step, we

Table 1: Correspondence between the two-competitor model [1] and the active inference account of induced realities.

Two-competitor model	Active inference account
Actual reality ar, induced reality ir as orthogonal axes	Two values of a hidden context $c \in \{\text{ar}, \text{ir}\}$ that selects the likelihood mapping and the transition dynamics of a hierarchical generative model, Eq. (4)
Degree of attention φ_a	Posterior over the context, $q(c = \text{ir}) = \sigma(\lambda)$ with $\lambda = \ln p(\text{ir})/p(\text{ar}) + F_{\text{ar}} - F_{\text{ir}}$, Eq. (5); φ_a is the angle of the precision allocation that drives λ
Magnitude m_a (total allocation)	Total precision deployed across both realities; its maintenance cost is attentional effort and sensory strain
Exogenous weights w^{exo}, map $\Phi_{\boldsymbol{p}}$	Entrenched likelihood precision of sensory events (abrupt onsets, looming, faces), part of the evolutionarily fixed core of the model
Endogenous weights w^{endo}, map $\Psi_{\boldsymbol{p}}$	Precision assigned to the streams that matter for the selected policies, set by expected free energy, Eq. (6)
Transfer function $\Omega_{\alpha,\beta,\gamma}$	Sigmoidal saturation of precision into a bounded resource allocation; exogenous priority as higher gain of core precisions
External vs. internal attention	Level of the hierarchy at which precision is deployed: sensorimotor contingencies (external) vs. scene, narrative, and schema states (internal)
Motivation, reward, effort (fatigue module)	Pragmatic value (expected preference fulfillment) and epistemic value (expected information gain) of the policies that act in each reality, Eq. (3); effort as the cost of precision
Affect	Valence as the rate of change of free energy; arousal as the interoceptive correlate of precision changes
Exogenous override (slamming door)	High-precision prediction error in an ar-specific stream that the induced model cannot absorb; Bayesian model switching with precision-dependent rate
Stimuli that "fit together"	Outputs of a single generative model in which action has predictable consequences (valid sensorimotor contingencies, counterfactual richness, plausible dynamics)
Multisensory integration	Shared hidden causes across modalities in $p(\boldsymbol{o} \mid \boldsymbol{s}, c)$; asynchrony and semantic incongruency as prediction errors
Closed-loop control of immersion	Online estimation of $q(c = \text{ir})$ from physiological markers of precision and prediction error, and adjustment of the stimuli that determine F_{ir} and $G(\xi)$

used a simplified dynamic update of the context log-odds λ. The context log-odds express the relative probability of the induced versus the actual reality on a logarithmic scale, and were updated by accumulating the difference in precision-weighted sensory evidence between the two realities. This accumulation was leaky, meaning that earlier evidence gradually lost its influence, and included a prior term informed by the posterior probabilities of the current policies. The update rate was proportional to the total precision, so that more strongly weighted evidence could lead to faster changes in the inferred context. The two-competitor angle φ_a was recovered from the same precisions through the original transfer functions and parameters of Strauss et al. [1]. Valence was estimated from the smoothed negative rate of change of a simplified free energy measure. This proxy combined the expected free energy of the policy mixture with a term representing unexpected exogenous events in the reality that is currently considered less likely. The update equations and parameters are given in Appendix A. The code that generates Figures 2 and 3 can be downloaded together with the Induced Reality Simulator, a browser application in which all parameters and events of the model can be varied and the resulting time courses inspected directly; the link is given in the statement on data and code availability at the end of the article.

Figure 2 shows the result. The account reproduces the qualitative behavior of the original simulation: a stable, high degree of immersion in the rewarding scenario with a transient break at the door, and a progressive loss of immersion in the fatiguing scenario with a deeper break and an incomplete recovery. The two-competitor angle φ_a, computed from the same precisions, follows $q(c = \mathrm{ir})$ closely but not identically: φ_a is an instantaneous function of the precisions and therefore responds without delay to the exogenous transient, whereas $q(c = \mathrm{ir})$ must reaccumulate evidence after the transient has passed, so that the two quantities dissociate briefly around every break. The simulation also makes explicit what the original model left implicit: in scenario B the loss of immersion is caused by the collapse of pragmatic and epistemic value, which withdraws endogenous precision from the induced reality before any exogenous event occurs, and the door only accelerates a switch that the expected free energy had already prepared.

Figure 3 shows the steady-state degree of immersion as a function of the two differentials that enter λ, the difference in precision-weighted sensory evidence and the difference in expected free energy between the two realities. The map organizes the situations that the term induced reality covers. Immersive virtual reality with a motivating task occupies the upper right, where sensory and motivational evidence agree. A novel or a daydream occupies the upper left, where the sensory evidence favors the actual reality but the policies lie in the induced one; immersion here is real but bounded, and it depends on the endogenous precision of the actual reality's streams being held at its floor, which is why reading is so easily disrupted. A high-fidelity system without a motivating task occupies the lower right and produces only moderate immersion, a situation familiar to anyone who has demonstrated a headset to an uninterested visitor. A break in immersion occupies the lower left. Augmented and mixed reality,

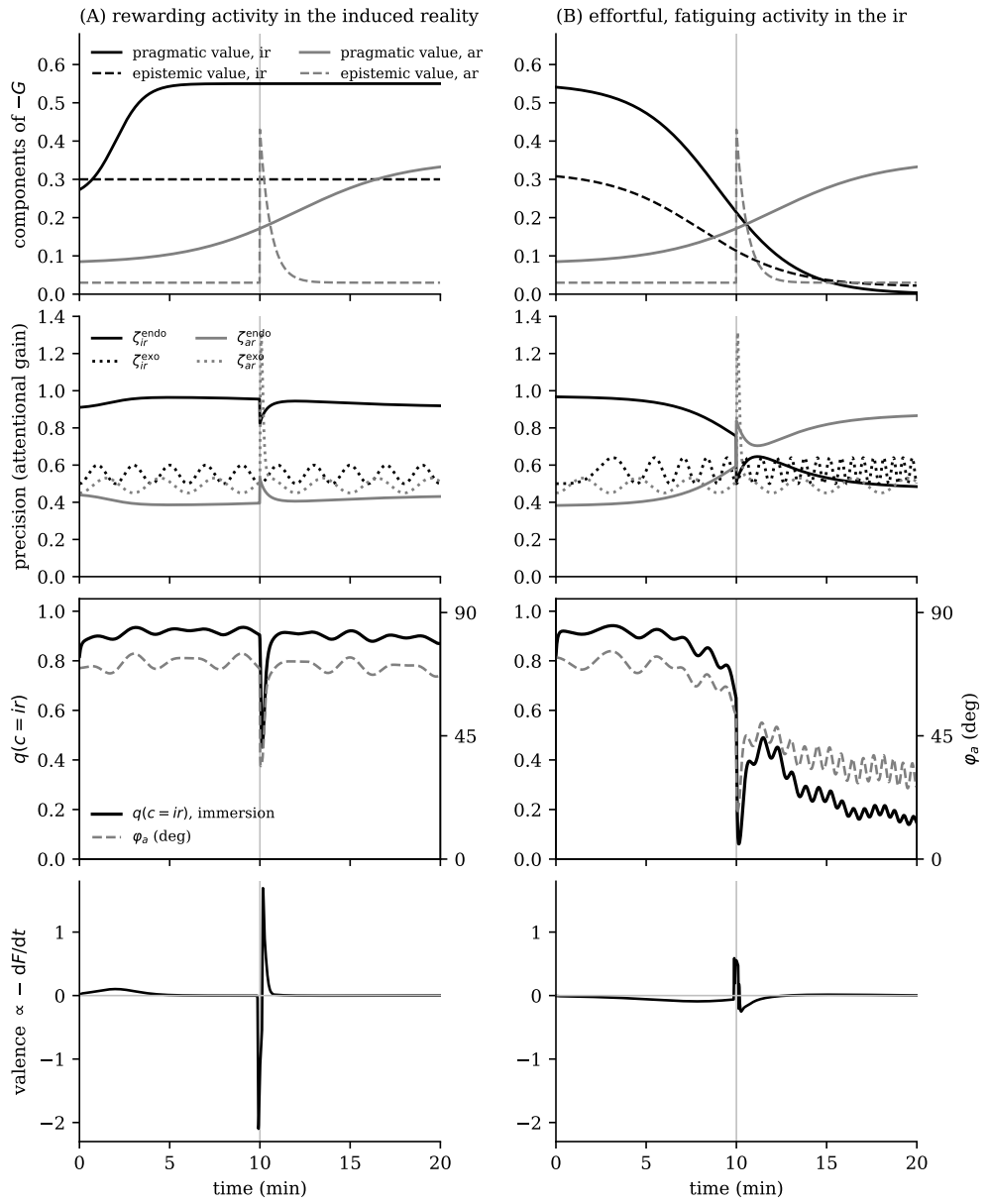


Figure 2: Active inference reproduction of the two scenarios of the two-competitor model. Left column: a rewarding activity in the induced reality (ir). Right column: an effortful, fatiguing activity. A slamming door in the actual reality (ar) occurs at $t = 10$ min (vertical line). First row: pragmatic and epistemic value of the policies that act in each reality, that is, the components of the negative expected free energy $-G$. Second row: endogenous precisions, which follow the policy posterior, and exogenous precisions, which follow the entrenched likelihood precision of sensory events; the door is a transient of $\zeta_{\mathrm{ar}}^{\mathrm{exo}}$. Third row: the degree of immersion $q(c = \mathrm{ir})$ (solid) and the two-competitor angle φ_a recovered through the original transfer functions (dashed). Fourth row: valence as the negative rate of change of the free energy proxy. In scenario A, immersion is high and stable, the door produces a fast break and a slower recovery, and valence is mildly positive while the pragmatic value of the induced reality rises and neutral thereafter, apart from the startle transient at the door. In scenario B, the collapse of pragmatic and epistemic value withdraws endogenous precision from the induced reality, immersion declines before the door, the break is deeper and the recovery incomplete, and valence is negative until the induced reality is abandoned. Compare with Figure 3 of Strauss et al. [1].

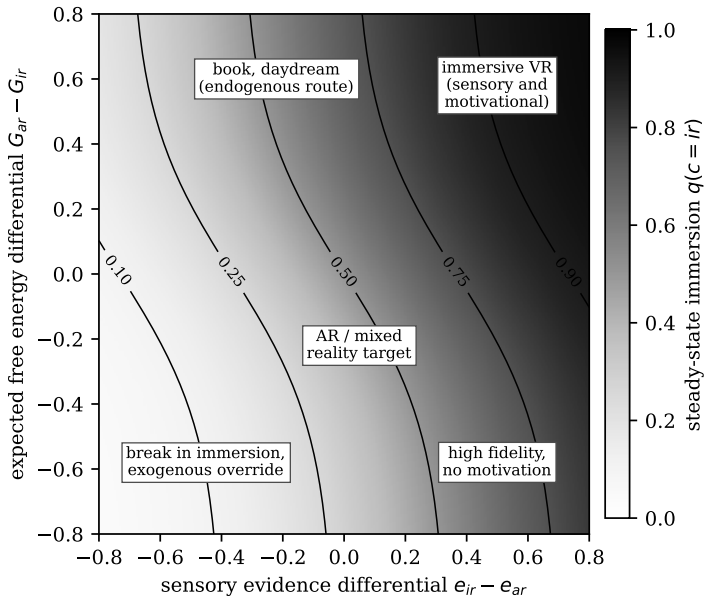


Figure 3: Steady-state degree of immersion $q(c = \mathrm{ir})$ as a function of the difference in precision-weighted sensory evidence between the two realities (horizontal axis) and the difference in expected free energy between the policies that act in the actual and in the induced reality (vertical axis), for the parameters of the simulation. The two differentials are treated as independent arguments of the steady-state log-odds, although in the full model the endogenous component of the sensory evidence itself depends on the policy posterior. The map organizes the situations covered by the term induced reality: immersive virtual reality combines sensory and motivational evidence; a book or a daydream relies on the endogenous route with weak or negative sensory evidence; a high-fidelity system without motivation and a break in immersion occupy the lower half; and augmented and mixed reality target an intermediate, stable superposition rather than a win.

finally, target the center of the map: their purpose is a stable superposition of the two realities rather than a win of either, so that the design goal is not $\varphi_a \to \pi/2$ but a stable intermediate $q(c)$ in which the augmenting content is explained as part of the actual reality. The map is also the quantitative form of the hypothesis that attention is necessary but not sufficient for immersion [1]: precision determines which streams count as evidence, but only the expected free energy of policies that lie in the induced reality turns attended evidence into belief, and the lower right of the map shows how little sensory evidence achieves on its own. We return to this point in Section 5.2.

5 Implications

5.1 Assessment of immersion

Subjective assessments of immersion and presence commonly rely on self-report questionnaires administered after the immersive experience, such as the Presence Questionnaire (PQ), the Immersive Experience Questionnaire for Film and TV (Film IEQ) or the Immersive Experience

Questionnaire (IEQ) [127–129]. While these measures are useful for capturing participants' conscious experience, they are mostly experience-specific and inherently retrospective and may therefore be influenced by recall biases, demand characteristics, and individual differences in the interpretation of questionnaire items. Moreover, post-hoc questionnaires provide limited temporal resolution and may fail to capture brief fluctuations or breaks in immersion, whereas continuous or in-situ self-reports can themselves interfere with the immersive experience [27].

In the two-competitor model we argued that the objective assessment of immersion should combine measures of the attentional focus vector's magnitude with measures of its direction, and that the latter requires the co-registration of events in both realities with evoked and event-related electrophysiological responses [1]. The active inference account sharpens this program by linking physiological measures to the precision allocation, prediction errors and the competition between actual and induced realities.

Event-related responses to violations of regularity, the mismatch negativity and the P3a [130, 131], are established correlates of precision-weighted prediction error [55, 56]. Probing both realities with matched irregular events (a deviant tone in the room, a deviant event in the virtual scene) and comparing the event-related responses may provide an indirect index of relative precision allocation, provided that differences in prediction error and sensory processing are appropriately controlled. The N1 attention effect and other sensory gain effects [132] may provide complementary indices of relative precision allocation at the level of early sensory processing. Our cross-modal semantic incongruency response to mismatching tactile and virtual visual experiences [24] and the prediction error signatures of implausible visuo-haptic interactions reported by Gehrke, Gramann, and colleagues [23, 25, 26] index sensory prediction errors under the induced model, which contribute to the term F_{ir} in Eq. (5).

Alpha power, which we and others have associated with internal attention and with the gating of sensory input [1, 48], indexes the attenuation of precision assigned to external sensory streams and is therefore particularly relevant for assessing internally maintained immersion, such as narrative absorption. The electromyogram of the vestigial auriculomotor system deserves a separate comment, because it gives access to both components that the account requires for immersion. The extrinsic periauricular muscles indicate the direction of exogenous and endogenous auditory attention [102, 103], raise their tonic activity during sustained selective listening [133], index listening effort [104], and show a brief suppression of the superior auricular muscle during reflexive orienting [134], so that their activity tracks the allocation of precision between the two realities and the cost of maintaining it. The intrinsic muscles tragicus and antitragicus, in turn, increase their activity to positively and to negatively arousing audiovisual events and correlate with subjective arousal [106], and the postauricular reflex to a probe sound is larger after a fearful than after a neutral face in far peripheral vision [105]. The same muscle system therefore also reads the affective side of the response, arousal, which the account relates to changes in precision (Section 3.3); the valence-related activity of the corrugator and zygomaticus muscles [135] adds the second affective dimension, which the account reads as

the rate of change of free energy. We have argued elsewhere that peripheral channels of this kind resolve affective processing at the timescale of the events that evoke it, together with its moment-to-moment coupling to attention and effort, which neither an observer nor self-report can resolve without interrupting the state that is to be measured [136]. Immersion, which requires attention and affective resonance together, is a case in point. The same work also marks the limits of such channels. Auricular activity is motor activity that is shaped by stimulus features, attention, and effort as well as by arousal, so that a rise is compatible with a shift of attention, with effort, and with excitement, and neither autonomic measures nor facial movements carry a fingerprint of emotion categories [137, 138]. In the present account this ambiguity is the expected form of the data rather than a defect, because attention, effort, and arousal are consequences of the same precision dynamics, and the markers are to be read as evidence in the estimation problem of Section 5.3, as a dimensional state with its uncertainty, rather than as labels. Together with skin conductance and heart rate [46, 139], the auricular electromyogram might then provide an estimate of the interoceptive precision changes that accompany a break in immersion, and it might separate a break that is only an orienting response from one that is also an affective event. The virtual presence counter of Slater and Steed [49] can be interpreted as a count of model switches, while the recovery time after each break, which the active inference account predicts to depend on endogenous precision, provides an additional dynamic indicator of the depth of immersion. Beyond the number of breaks, their temporal dynamics may provide further insight into the processes underlying immersion: while a high-precision prediction error can rapidly disrupt immersion, recovery requires the re-accumulation of evidence for the induced reality, a process that depends in part on endogenous precision allocation and might therefore serve as a marker.

Immersion is inherently individual and should be considered as such: even a realistic and internally coherent virtual environment does not necessarily produce the same degree of immersion across individuals. From an active inference perspective, Parvizi-Wayne [73] argues that attention is shaped by expectations and preferences that extend beyond an immediate goal. Accordingly, the allocation of attention toward an induced reality may depend on personal experiences, long-term preferences, and emotional relevance. Bouizegarene et al. [140] describe narratives as integrative structures at higher levels of generative models, organizing predictions about temporally and causally related events. These predictions draw on individual experiences and culturally shared expectations, while new events can in turn modify existing beliefs. Consequently, the same story may be interpreted differently depending on the expectations and beliefs a person brings to it. Experimental evidence for this comes from several studies examining differences in neural responses to narratives. Presenting the same ambiguous story with different interpretive contexts beforehand led to divergent fMRI responses between groups, while responses were more similar among individuals who shared the same interpretation [141]. Importantly, such differences are not limited to externally provided interpretive cues. Finn et al. [142] found that individual differences in trait paranoia were associated with distinct neural re-

sponses to the same narrative. More recent findings indicate that belief and disbelief influence neural synchronization during narrative processing [143], while greater similarity in narrative understanding is associated with more similar neural responses [144]. These differences in prior knowledge and interpretation may also influence the degree of immersion. For example, prior knowledge or experience relevant to a story's themes has been associated with greater narrative transportation [44]. Individuals also differ in their general tendency toward absorption, a disposition assessed by the Tellegen Absorption Scale and associated with hypnotic susceptibility and aspects of openness to experience [145].

Three experimental paradigms follow. First, a graded violation paradigm, in which the sensorimotor contingencies of the induced reality are parametrically degraded (latency, tracking noise, visuo-haptic offset) while prediction error markers and the reported break rate are recorded, tests the prediction that increasing violations of sensorimotor contingencies are accompanied by stronger physiological prediction-error responses and an increased rate of breaks in presence, and that the tolerance is larger for mantle contingencies than for core contingencies. Second, a dual-probe paradigm, in which matched exogenous probes are delivered in both realities at unpredictable times, tests the prediction that the ratio of the evoked responses provides an index of φ_a and that the auriculomotor response to an ar probe is larger when the user is more immersed, because the probe is then more surprising under the current model. Because the postauricular reflex is also modulated by the affective content of the preceding event [105], the probes must be affectively neutral and matched across the two realities, and the tonic activity of the intrinsic muscles provides a covariate for arousal from the same recording. Third, an epistemic value paradigm, in which the difficulty or the novelty of the task in the induced reality is manipulated to produce boredom, flow, and overload, tests the prediction that immersion follows the expected information gain and that effort markers [3, 104] dissociate from immersion markers in the overload condition, where effort is high and immersion declines.

5.2 Design principles for immersive technology

Abrash's dictum that a virtual reality only has to provide the right inputs to satisfy the relevant senses [9, 10] becomes, in the present account, a hierarchy of requirements ordered by precision. The requirements that Abrash and colleagues derived for presence a decade ago [10, 11], a wide field of view, low persistence, high refresh rate, precise tracking, and low motion-to-photon latency, all concern the fast, entrenched core of the perceptual system: the sensorimotor contingencies between self-motion and optic flow, the vestibular-visual coupling, and the temporal precision of the visual system. These contingencies carry high, non-negotiable precision, so that their violation produces prediction errors that no narrative can absorb. Requirements that concern the mantle, such as the fidelity of textures, the realism of avatars, or the naturalness of a particular interaction metaphor, are negotiable and can be traded against cost, because the corresponding precisions are learned and can be lowered by experience. The

recent research directions of Reality Labs that Abrash described, varifocal displays that satisfy the accommodation-vergence contingency, near-retinal resolution, wide field of view, and electromyographic wristbands for low-friction action [9], are all attempts to satisfy core contingencies, and the account explains why they matter more than photorealism.

The account also gives a reason for two design intuitions that are widely shared but rarely justified. First, multisensory congruency matters more than unisensory fidelity [1, 146, 147]: congruent streams are explained by a shared hidden cause and therefore reduce F_{ir} jointly, whereas an incongruent stream, however good in isolation, adds a prediction error. Second, narrative coherence matters even in sensory-driven immersion, because the deep levels of the induced model are the ones that maintain the endogenous precision of its streams. A world that responds to the user, in which events have causes and in which the user's actions have consequences, keeps the epistemic value of the policies that act in it high. Adaptive difficulty, which game designers have long used to keep players in flow, is in these terms the regulation of epistemic value, and it can be made explicit by estimating the rate at which the user resolves uncertainty and adjusting the task accordingly.

Augmented and mixed reality present a different design goal, which the map of Figure 3 makes explicit. Here the augmenting content must be explained as part of the actual reality: it must respect the sensorimotor contingencies of the room, be occluded by real objects, cast shadows, and, above all, not compete with the actual reality's streams for precision when the actual reality matters, as it does when walking through traffic or operating a machine. The design goal is a stable, intermediate $q(c)$, a superposition in which the augmenting content inherits the transparency of the actual reality rather than competing with it. Abrash's description of contextual artificial intelligence in augmented reality glasses, which infers what the user needs from the situation and delivers it with minimal friction [9], is a description of a system that manages the precision of its own outputs so as to reduce, rather than to increase, the free energy of the user's model of the actual reality. This connects to our neuroergonomic proposals for attention-aware modality selection in digital operating rooms and industrial settings [5, 148**?**]: the system should route critical information to the modality and the location in the scale space where precision is currently available, and it should refrain from adding streams that would have to be suppressed.

5.3 Closed-loop control of immersion

In the two-competitor model we proposed closed-loop systems that use physiological measures to maintain immersion at an intended level, for instance by sending an appropriate stimulus when a drift of attention away from the induced reality is detected [1]. The active inference account turns this proposal into an estimation and control problem. The state to be estimated is $q(c = \mathrm{ir})$, together with the precisions and the expected free energy terms that drive it; the observations are the physiological markers of Section 5.1; and the control variables are the stimulus

properties that determine F_{ir} (coherence, latency, congruency), the events that determine exogenous precision (salient events in the induced reality), and the task properties that determine pragmatic and epistemic value (reward, difficulty, novelty). A controller that keeps λ in a target range by acting on these variables is a neuroadaptive system in the sense of our work on empathetic, context-sensitive human-technology interaction [5, 148]. We have recently specified such a loop for affect-aware systems in general: a generative model of affect grounded in active inference carries a state estimate across time, predicts the effect of the system's own actions on that state, and selects actions by expected free energy, under bounded loop gain and with human oversight [136]. An immersion controller is a special case in which the estimated state includes $q(c = \text{ir})$ and the objective is a target level of immersion, and it inherits the conditions under which such a loop is steerable rather than manipulative, in particular an objective set by the user and explicit, revocable consent for every sensing channel. The auricular channel of Section 5.1 is attractive for this purpose because electrodes in and around the ear, which can be integrated into headsets and earbuds, deliver the direction of attention, effort, and arousal from one site [103, 136]. The estimation side of the loop is what we have proposed, for neuroergonomic design in general, as a surrogate human digital twin: a generative model of the latent processes behind embodied perception and cognition that updates its beliefs from physiological and behavioral channels weighted by their reliability, and that permits counterfactual tests in which priors or inputs are changed [149]. With the model of Section 3 as its computational backbone, such a twin estimates $q(c = \text{ir})$ together with the precisions and the expected free energy terms that drive it, and it allows a design variant, an added latency, an incongruent stream, a change in task difficulty, to be tested on the twin before it is tested on people. Two features of the account are relevant for such a controller. First, the asymmetry between the fast break and the slow recovery implies that prevention is more effective than repair: a controller should reduce the probability of high-precision ar events (by noise management and by scheduling) rather than react to them. Second, the collapse of epistemic value in the overload and the boredom regimes implies that the controller should monitor the rate of uncertainty resolution, for which effort markers and performance are the natural observations, and adjust difficulty before immersion declines. For therapeutic applications, in which the level of immersion or of arousal is itself the therapeutic variable, as in exposure therapy or in rehabilitation [1, 21, 150], the same controller can be used to hold $q(c = \text{ir})$ or the interoceptive precision at a prescribed level.

5.4 Neuroergonomics and the Anthropocene

We have argued that the technological environments of the Anthropocene impose sensory conflict and sensory strain on a perceptual system that evolved for different environments [5, 6]. Immersive technologies are the most extreme instance of such environments, because they replace, rather than merely alter, the sensory statistics to which the core is tuned. The present account locates the two problems in the same variables. Sensory conflict is free energy at the

core that neither perception nor action can reduce; it is what makes an induced reality sickening. Sensory strain is the sustained cost of maintaining the precision allocation that keeps λ positive against a competing reality; it is what makes an induced reality tiring. Both are increased by every stream that the induced model must suppress rather than explain, and both are reduced by design that respects the core. The account therefore supports the neuroergonomic principle we have proposed, that technology design should start from the sensory statistics to which the core is tuned [5], and it adds a prediction: an induced reality that explains the bodily streams (through embodiment, through consistent self-motion, through interoceptive plausibility) is more immersive and less straining than one that merely suppresses them.

5.5 Reality monitoring

Presence, in the definition of Section 3.1, is the absence of prediction error at the context level. Slater et al. noted that a perfect virtual reality, indistinguishable from the physical world, would abolish presence as a distinct experience because the participant would not know that a virtual reality was in use [34]. In the present terms, a perfect induced reality is one in which $F_{\mathrm{ir}} \leq F_{\mathrm{ar}}$, so that no evidence remains that could restore $q(c = \mathrm{ar})$. Reality monitoring, the capacity to distinguish perception from imagination and mediated from unmediated experience, depends on such evidence [22, 81, 82]. One design consequence follows: a route back to the actual reality remains available at low cost only if there is a class of events in the actual reality whose precision remains sufficiently high during immersion.

6 Limitations and outlook

The account presented here is a theoretical framework, and several of its elements are simplifications that future work must address. First, we have treated the context as a binary variable. The examples of Section 2.1 do not require more, because a cue of any physical origin is assigned to one of the two realities by its coherence with the scene. Real situations do, however, involve nested realities (a virtual meeting inside a physical office, a story within a game, a memory within a daydream) in which more than two explanations compete and the context is itself hierarchical. The extension to a set of contexts, or to a continuous context space, is immediate in principle, since Eq. (5) generalizes to a softmax over any number of sub-models, but the empirical question of how many contexts the brain entertains at once, and at what cost, is open. Second, the simulation of Section 4.2 specified the pragmatic and epistemic values as time courses. A full implementation would compute them from a generative model of the task in the induced reality, using the discrete-state formulation of active inference [12, 54, 151], and would let learning change the model over sessions. This is the natural next step, and it would allow quantitative predictions for the graded violation and epistemic value paradigms proposed above. Third, we have said little about the neuronal implementation. The hierarchical message

passing of active inference has been mapped onto cortical microcircuits and onto the thalamocortical regulation of precision [56, 57, 65, 152], and the context inference of Eq. (5) presumably involves the same medial prefrontal and hippocampal circuits that support scene construction and reality monitoring [81, 153], but a specific proposal must await data from paradigms of the kind we have outlined. Fourth, the psychophysiological markers we have listed are correlates of precision and prediction error in general, not of the context inference in particular; their specificity must be established by co-registration with the events of both realities, as we argued in the two-competitor model. Fifth, individual differences, which we have located in the mantle [5], enter the account through the learned precisions and the prior preferences and are therefore expected to be large; the account predicts, for instance, that individuals with high interoceptive precision are less immersible by exteroceptive means and more immersible by embodiment, a prediction that can be tested. Finally, we have restricted ourselves to a single agent. Social presence, the sense of being with another person, which Abrash regards as the decisive application of immersive technology [9], requires a model in which the other agent's states are inferred and in which the induced reality must support the sensorimotor contingencies of interaction, gaze, turn-taking, and touch; the framework extends to this case, but we have not developed it here.

7 Conclusions

We have proposed that immersion is inference. An induced reality is a candidate generative model that competes with the model of the actual reality for the explanation of a shared sensorium and for the control of action; the degree of immersion is the posterior probability that the induced reality is the current context; presence is the transparency of the winning model; and a break in immersion is a Bayesian switch driven by a high-precision prediction error that the induced model cannot absorb. The exogenous and endogenous components of the two-competitor model are two routes of precision allocation, one set by the evolutionarily entrenched core of the perceptual system and one set by the expected free energy of the agent's policies, and the affective and motivational dynamics that the two-competitor model captured with a fatigue module are the pragmatic and epistemic value of remaining in the induced reality. The account recovers the quantities of the two-competitor model, reproduces its scenarios, and adds a perceptual variable, the context posterior, that the attentional focus vector drives but does not equal. This is the formal content of the hypothesis that attention is necessary but not sufficient for immersion: attention sets the evidence, and the expected free energy of the policies in the induced reality decides whether the evidence becomes belief. The account connects presence, embodiment, cybersickness, and narrative absorption to a single quantity; it specifies what the psychophysiological markers of immersion measure; and it yields a design hierarchy in which the core of perception is non-negotiable and the mantle can be trained. The engineering intuition that a virtual reality only has to satisfy the senses, not replicate the world, is thereby given its neuroscientific form: the senses are satisfied when the world they are offered is the one that best

explains what they sense and best serves what the person is trying to do.

Author contributions

LR: Methodology, Investigation, Writing (original draft), Writing (review and editing). KLG: Writing (review and editing). JV: Conceptualization, Investigation, Methodology, Writing (review and editing). ALF: Conceptualization, Formal analysis, Investigation, Methodology, Writing (review and editing). FIC-S: Conceptualization, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing (review and editing). DJS: Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualization, Writing (original draft), Writing (review and editing), Funding acquisition, Supervision.

Funding

This work was supported by the German Federal Ministry of Research, Technology, and Space (BMFTR) through the grant 13FH050KX1 (Deep Immersion Lab Saar), by the State of Saarland through the Saarland Transformationsfonds (SaarNex, Trafo 25/04), and by the European Union (European Regional Development Fund, ERDF) and Saarland through the Center for Digital Neurotechnologies Saar (EFRE-HS-0000835) and Multi-Immerse (EFRE-HS-0000832) as well as the collaborative project Saarland 5.0 (Human-Centered Digital Neurotechnologies for a Healthy Saarland 5.0).

Declaration of competing interest

The authors declare no competing interests.

Data and code availability

No empirical data were used. The Python code that generates Figures 2 and 3 and the Induced Reality Simulator, a browser application in HTML5/JavaScript that runs without installation and in which all parameters and events of the model can be varied, can be downloaded from https://cloud.hiz-saarland.de/s/S79mStn5TagwpTC. The package also contains the settings of the two scenarios of Section 4.2. We will make the code publicly available on GitHub after acceptance.

Generative AI statement

The author(s) declared that Generative AI was used in the creation of this manuscript. During the preparation of this work the authors used a large language model to assist with literature triage and language editing. The authors reviewed and edited all content and take full responsibility for the publication.

A Simulation details

The simulation of Section 4.2 uses a discrete time step of $\Delta t = 1$ s over 20 min, that is, $T = 1200$ steps. For each reality $c \in \{\mathrm{ar}, \mathrm{ir}\}$, the pragmatic value $r_c(t)$ and the epistemic value $\varepsilon_c(t)$ of the policies acting in c are specified as time courses (Figure 2, first row), and the expected free energy of these policies is calculated as $G_c(t) = -\big(r_c(t) + \varepsilon_c(t)\big)$. The policy posterior is

$$q_t(\xi_{\mathrm{ir}}) = \sigma\big(\gamma_\xi \, (G_{\mathrm{ar}}(t) - G_{\mathrm{ir}}(t))\big), \qquad q_t(\xi_{\mathrm{ar}}) = 1 - q_t(\xi_{\mathrm{ir}}), \tag{7}$$

with $\gamma_\xi = 4$. Endogenous precisions follow Eq. (6) with modeling assumptions $\zeta_0 = 0.35$ and $\zeta_{\max} = 1$. Exogenous precisions $\zeta_c^{\mathrm{exo}}(t)$ are oscillatory functions with a baseline of 0.5 (ir) and 0.45 (ar); the slamming door adds, from $t = 10$ min on, a causal alpha-function transient $0.8\,x\,e^{1-x}$ with $x = (t - t_{\mathrm{door}})/4$ s (peak after 4 s) to $\zeta_{\mathrm{ar}}^{\mathrm{exo}}$ and an exponentially decaying epistemic value of amplitude 0.4 and time constant 35 s to $\varepsilon_{\mathrm{ar}}$; the brief transient and the slower decay of the epistemic value are what produce the fast break and the slower recovery in Figure 2, together with the precision-dependent update rate below. The precision-weighted sensory evidence of each reality is $e_c(t) = \big(\zeta_c^{\mathrm{endo}}(t) + \zeta_c^{\mathrm{exo}}(t)\big)\, s_c$, where s_c is the coherence of the reality's model with its streams once attended ($s_{\mathrm{ir}} = 0.90$, $s_{\mathrm{ar}} = 0.95$). The context log-odds relax toward a target that combines the evidence differential with a policy-informed prior,

$$\lambda^*(t) = \kappa\big(e_{\mathrm{ir}}(t) - e_{\mathrm{ar}}(t)\big) + \rho\,\big(2\,q_t(\xi_{\mathrm{ir}}) - 1\big), \qquad \frac{\mathrm{d}\lambda}{\mathrm{d}t} = \frac{Z(t)}{\tau_0}\,\big(\lambda^*(t) - \lambda(t)\big), \tag{8}$$

where $Z(t)$ is the sum of all four precisions, so that the rate of belief updating scales with the total precision, and $\kappa = 3$, $\rho = 1$, $\tau_0 = 15$ s. The degree of immersion is $q_t(c = \mathrm{ir}) = \sigma(\lambda(t))$. The two-competitor projections are computed from the same precisions with the original transfer functions and parameters, $a_c = \Omega_{0.58,7,5}(\zeta_c^{\mathrm{exo}}) + \Omega_{0.42,9,5}(\zeta_c^{\mathrm{endo}})$, and $\varphi_a = \tan^{-1}(a_{\mathrm{ir}}/a_{\mathrm{ar}})$. The free energy proxy is

$$F(t) = \sum_{\xi} q_t(\xi)\, G_\xi(t) + q_t(c = \mathrm{ir})\,\big[\zeta_{\mathrm{ar}}^{\mathrm{exo}}(t) - \bar{\zeta}_{\mathrm{ar}}^{\mathrm{exo}}\big]_+ + q_t(c = \mathrm{ar})\,\big[\zeta_{\mathrm{ir}}^{\mathrm{exo}}(t) - \bar{\zeta}_{\mathrm{ir}}^{\mathrm{exo}}\big]_+, \tag{9}$$

where $[\cdot]_+$ denotes rectification and $\bar{\zeta}_c^{\mathrm{exo}}$ is the maximum of the expected oscillatory exogenous precision (0.53 for ar, 0.64 for ir), so that only unexpected exogenous events in the

reality one does not currently believe in contribute unexplained prediction error. Valence is $-\mathrm{d}F/\mathrm{d}t$ smoothed with a 15 s moving average. All quantities are dimensionless except time; the log-odds are initialized at $\lambda(0) = 1.5$, valence is expressed per minute, and the epistemic values have small floors (0.03 for ar, 0.02 for ir in scenario B). The pragmatic values r_c are affine proxies of the expected log-probability of preferred outcomes rather than log-probabilities themselves. The Python code is `simulate_immersion.py`, part of the code package given in the statement on data and code availability; the settings of scenarios A and B are included as `settings_preset_A.json` and `settings_preset_B.json` and can be loaded into the browser application.

References

[1] D. J. Strauss, A. L. Francis, J. Vibell, and F. I. Corona–Strauss. The role of attention in immersion: The two–competitor model. *Brain Research Bulletin*, 210:110923, 2024. doi: 10.1016/j.brainresbull.2024.110923.

[2] M. M. Chun, J. D. Golomb, and N. B. Turk-Browne. A taxonomy of external and internal attention. *Annual Review of Psychology*, 62:73–101, 2011. doi: 10.1146/annurev.psych.093008.100427.

[3] D. J. Strauss and A. J. Francis. Toward taxonomic model of attention in effortful listening. *Cogn Affect Behav Neurosci.*, 17:809–825, 2017. doi: 10.3758/s13415-017-0513-0.

[4] E. N. Schneider, C. Bernarding, A. L. Francis, B. W. Y. Hornsby, and D. J. Strauss. A quantitative model of listening related fatigue. In *Neural Engineering (NER), 9th International IEEE/EMBS Conference on*, pages 619–622, 2019. doi: 10.1109/NER.2019.8717046.

[5] F. I. Corona-Strauss, J. Vibell, A. L. Francis, M. Lehser, S. M. Markert, and D. J. Strauss. Reframing neuroergonomics in an evolutionary and active inference context. *Frontiers in Neuroergonomics*, 7:1796721, 2026. doi: 10.3389/fnrgo.2026.1796721.

[6] D. J. Strauss, A. L. Francis, Z. Schäfer, M. Latzel, F. I. Corona–Strauss, and S. Launer. Understanding speech in "noise" or free energy minimization in the soundscapes of the anthropocene. *Frontiers in Neuroscience*, 19, 2025. doi: 10.3389/fnins.2025.1534425.

[7] K. J. Friston. The free-energy principle: a unified brain theory? *Nat Rev Neurosci.*, 11:127–138, 2010. doi: 10.1038/nrn2787.

[8] T. Parr, G. Pezzulo, and K. J. Friston. *Active inference: the free energy principle in mind, brain, and behavior*. MIT Press, 2022. doi: 10.7551/mitpress/12441.001.0001.

[9] M. Abrash. Reality Labs research: A look back, a look forward. Paul G. Allen School Distinguished Lecture, University of Washington, Seattle, WA, December 4, 2025. https://www.youtube.com/watch?v=V5AwlNbAvCo (last accessed on 30.10.2026, 09:30).

[10] M. Abrash. Why virtual reality will matter to you. Keynote at the Facebook F8 Developer Conference, San Francisco, CA, March 26, 2015. https://www.youtube.com/watch?v=UDu-cnXI8E8 (last accessed on 30.10.2026, 09:30).

[11] M. Abrash. What VR could, should, and almost certainly will be within two years. Talk at Steam Dev Days, Seattle, WA, January 2014; slides and notes. https://media.steampowered.com/apps/abrashblog/Abrash%20Dev%20Days%202014.pdf (last accessed on 30.10.2026, 09:30).

[12] K. Friston, T. FitzGerald, F. Rigoli, P. Schwartenbeck, and G. Pezzulo. Active inference: A process theory. *Neural Computation*, 29:1–49, 2017. doi: 10.1162/NECO_a_00912.

[13] A. K. Seth, K. Suzuki, and H. D. Critchley. An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology*, 2:395, 2012. doi: 10.3389/fpsyg.2011.00395.

[14] A. K. Seth. Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17(11):565–573, 2013. doi: 10.1016/j.tics.2013.09.007.

[15] A. K. Seth. A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive Neuroscience*, 5:97–118, 2014. doi: 10.1080/17588928.2013.877880.

[16] J. K. O'Regan and A. Noë. A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24:939–973, 2001. doi: 10.1017/S0140525X01000115.

[17] J. Limanowski and K. Friston. 'Seeing the dark': Grounding phenomenal transparency and opacity in precision estimation for active inference. *Frontiers in Psychology*, 9:643, 2018. doi: 10.3389/fpsyg.2018.00643.

[18] T. K. Metzinger. Why is virtual reality interesting for philosophers? *Frontiers in Robotics and AI*, 5:101, 2018. doi: 10.3389/frobt.2018.00101.

[19] M. Samad, A. J. Chung, and L. Shams. Perception of body ownership is driven by Bayesian sensory inference. *PLoS ONE*, 10:e0117178, 2015. doi: 10.1371/journal.pone.0117178.

[20] M. A. J. Apps and M. Tsakiris. The free-energy self: A predictive coding account of self-recognition. *Neuroscience & Biobehavioral Reviews*, 41:85–97, 2014. doi: 10.1016/j.neubiorev.2013.01.029.

[21] G. Riva, B. K. Wiederhold, and F. Mantovani. Neuroscience of virtual reality: From virtual exposure to embodied medicine. *Cyberpsychology, Behavior, and Social Networking*, 22:82–96, 2019. doi: 10.1089/cyber.2017.29099.gri.

[22] O. Singleton and A. Fotopoulou. Presence is reality: Rethinking virtual and real-world consciousness. *Perspectives on Psychological Science*, 2026. doi: 10.1177/17456916251414030.

[23] L. Gehrke, P. Lopes, M. Klug, S. Akman, and K. Gramann. Neural sources of prediction errors detect unrealistic VR interactions. *Journal of Neural Engineering*, 19:036002, 2022. doi: 10.1088/1741-2552/ac69bc.

[24] C. Lehser, S. A. Hillyard, and D. J. Strauss. Feeling senseless sensations: a crossmodal EEG study of mismatched tactile and visual experiences in virtual reality. *Journal of Neural Engineering*, 2024. doi: 10.1088/1741-2552/ad83f5.

[25] L. Gehrke, L. Terfurth, and K. Gramann. Moderating role of presence in EEG responses to visuo-haptic prediction error in virtual reality. arXiv:2510.23262, 2025.

[26] E. G. Center, M. Pouke, A. Nardi, L. Gehrke, K. Gramann, T. Ojala, and S. M. LaValle. Virtual body ownership affects the processing of sensorimotor contingencies and goal-oriented information extraction, but not semantics. *NeuroImage*, 327:121715, 2026. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2026.121715.

[27] S. Agrawal, A. Simon, S. Bech, K. Bærentsen, and S. Forchhammer. Defining immersion: Literature review and implications for research on immersive audiovisual experiences. *Journal of the Audio Engineering Society*, 68:404–417, 2020. doi: 10.17743/jaes.2020.0039.

[28] R. Skarbez, F. P. Brooks, Jr., and M. C. Whitton. A survey of presence and related concepts. *ACM Computing Surveys*, 50:96, 2017. doi: 10.1145/3134301.

[29] N. C. Nilsson, R. Nordahl, and S. Serafin. Immersion revisited: A review of existing definitions of immersion and their relation to different theories of presence. *Human Technology*, 12(2):108–134, 2016. doi: 10.17011/ht/urn.201611174652.

[30] M. Slater and S. Wilbur. A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 6:603–616, 1997. doi: 10.1162/pres.1997.6.6.603.

[31] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B*, 364:3549–3557, 2009. doi: 10.1098/rstb.2009.0138.

[32] M. Slater. A note on presence terminology. *Presence connect*, 3, 2003.

[33] M. Slater, D. Banakou, A. Beacco, J. Gallego, F. Macia-Varela, and R. Oliva. A separate reality: An update on place illusion and plausibility in virtual reality. *Frontiers in Virtual Reality*, 3:914392, 2022. doi: 10.3389/frvir.2022.914392.

[34] M. Slater and M. V. Sanchez-Vives. Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI*, 3:74, 2016. doi: 10.3389/frobt.2016.00074.

[35] M. Lombard and T. Ditton. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2), 1997. doi: 10.1111/j.1083-6101.1997.tb00072.x.

[36] W. Wirth, T. Hartmann, S. Böcking, P. Vorderer, C. Klimmt, H. Schramm, T. Saari, J. Laarni, N. Ravaja, F. R. Gouveia, F. Biocca, A. Sacau, L. Jäncke, T. Baumgartner, and P. Jäncke. A process model of the formation of spatial presence experiences. *Media Psychology*, 9:493–525, 2007. doi: 10.1080/15213260701283079.

[37] T. Hartmann, W. Wirth, P. Vorderer, C. Klimmt, H. Schramm, and S. Böcking. Spatial presence theory: State of the art and challenges ahead. In M. Lombard, F. Biocca, J. Freeman, W. IJsselsteijn, and R. J. Schaevitz, editors, *Immersed in Media: Telepresence Theory, Measurement and Technology*, pages 115–135. Springer, Cham, 2015. doi: 10.1007/978-3-319-10190-3_7.

[38] K. Kilteni, R. Groten, and M. Slater. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21:373–387, 2012. doi: 10.1162/PRES_a_00124.

[39] K. Kilteni, A. Maselli, K. P. Kording, and M. Slater. Over my fake body: Body ownership illusions for studying the multisensory basis of own-body perception. *Frontiers in Human Neuroscience*, 9:141, 2015. doi: 10.3389/fnhum.2015.00141.

[40] O. Blanke, M. Slater, and A. Serino. Behavioral, neural, and computational principles of bodily self-consciousness. *Neuron*, 88:145–166, 2015. doi: 10.1016/j.neuron.2015.09.029.

[41] M. Czikszentmihalyi. *Flow: The psychology of optimal experience*. New York: Harper & Row, 1990.

[42] J. Nakamura and M. Csikszentmihalyi. The concept of flow. In *Flow and the foundations of positive psychology*, pages 239–263. Springer, 2014. doi: 10.1007/978-94-017-9088-8_16.

[43] M. C. Green and T. C. Brock. The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, 79(5):701–721, 2000. doi: 10.1037/0022-3514.79.5.701.

[44] M. C. Green. Transportation into narrative worlds: The role of prior knowledge and perceived realism. *Discourse processes*, 38(2):247–266, 2004. doi: 10.1207/s15326950dp3802_5.

[45] J. J. Cummings and J. N. Bailenson. How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, 19:272–309, 2016. doi: 10.1080/15213269.2015.1015740.

[46] M. Meehan, B. Insko, M. Whitton, and F. P. Brooks, Jr. Physiological measures of presence in stressful virtual environments. *ACM Transactions on Graphics*, 21:645–652, 2002. doi: 10.1145/566654.566630.

[47] T. Baumgartner, L. Valko, M. Esslen, and L. Jäncke. Neural correlate of spatial presence in an arousing and noninteractive virtual reality: An EEG and psychophysiology study. *CyberPsychology & Behavior*, 9:30–45, 2006. doi: 10.1089/cpb.2006.9.30.

[48] S. E. Kober, J. Kurzmann, and C. Neuper. Cortical correlate of spatial presence in 2D and 3D interactive virtual reality: an eeg study. *International Journal of Psychophysiology*, 83(3):365–374, 2012. doi: 10.1016/j.ijpsycho.2011.12.003.

[49] M. Slater and A. Steed. A virtual presence counter. *Presence: Teleoperators and Virtual Environments*, 9:413–434, 2000. doi: 10.1162/105474600566925.

[50] S. Pradhan, S. Pradhan, S. Laraway, and S. Snycerski. "no laughing matter": The influence and necessity of the laugh track on humor and enjoyability in a comedic sitcom. *The Journal of Communication and Media Studies*, 6(4):1–14, 2021. doi: 10.18848/2470-9247/CGP/v06i04/1-14.

[51] P. Hauck and H. Hecht. Emotionally congruent music and text increase immersion and appraisal. *Plos one*, 18(1):e0280019, 2023. doi: 10.1371/journal.pone.0280019.

[52] M. Richardson, I. D. Gilchrist, and D. Stanton Fraser. Cross-sensory interaction in immersion: Critiquing the two competitor model. Position paper, Cross-Sensory Futures Workshop at ACM SIGCHI 2026, Barcelona; PsyArXiv preprint, 2026. doi: 10.31234/osf.io/f47ca_v1.

[53] A. M. Baranowski, R. Teichmann, and H. Hecht. Canned emotions: Effects of genre and audience reaction on emotions. *Art & Perception*, 5(3):312–336, 2017. doi: 10.1163/22134913-00002068.

[54] R. Smith, K. J. Friston, and C. J. Whyte. A step-by-step tutorial on active inference and its application to empirical data. *Journal of Mathematical Psychology*, 107, 2022. ISSN 0022-2496. doi: https://doi.org/10.1016/j.jmp.2021.102632.

[55] H. Feldman and K. J. Friston. Attention, uncertainty, and free-energy. *Frontiers in human neuroscience*, 4:215, 2010. doi: 10.3389/fnhum.2010.00215.

[56] A. M. Bastos, W. M. Usrey, R. A. Adams, G. R. Mangun, P. Fries, and K. J. Friston. Canonical microcircuits for predictive coding. *Neuron*, 76:695–711, 2012. doi: 10.1016/j.neuron.2012.10.038.

[57] R. Kanai, Y. Komura, S. Shipp, and K. Friston. Cerebral hierarchies: Predictive processing, precision and the pulvinar. *Philosophical Transactions of the Royal Society B*, 370: 20140169, 2015. doi: 10.1098/rstb.2014.0169.

[58] T. Parr and K. J. Friston. Uncertainty, epistemics and active inference. *Journal of the Royal Society Interface*, 14:20170376, 2017. doi: 10.1098/rsif.2017.0376.

[59] T. Parr and K. J. Friston. Attention or salience? *Current Opinion in Psychology*, 29:1–5, 2019. doi: 10.1016/j.copsyc.2018.10.006.

[60] K. Friston, F. Rigoli, D. Ognibene, C. Mathys, T. Fitzgerald, and G. Pezzulo. Active inference and epistemic value. *Cognitive Neuroscience*, 6:187–214, 2015. doi: 10.1080/17588928.2015.1020053.

[61] P. Schwartenbeck, J. Passecker, T. U. Hauser, T. H. B. FitzGerald, M. Kronbichler, and K. J. Friston. Computational mechanisms of curiosity and goal-directed exploration. *eLife*, 8:e41703, 2019. doi: 10.7554/eLife.41703.

[62] J. Kiverstein, M. Miller, and E. Rietveld. The feeling of grip: Novelty, error dynamics, and the predictive brain. *Synthese*, 196:2847–2869, 2019. doi: 10.1007/s11229-017-1583-9.

[63] K. J. Friston, R. Rosch, T. Parr, C. Price, and H. Bowman. Deep temporal models and active inference. *Neuroscience & Biobehavioral Reviews*, 77:388–402, 2017. doi: 10.1016/j.neubiorev.2017.04.009.

[64] G. Pezzulo, F. Rigoli, and K. J. Friston. Hierarchical active inference: A theory of motivated control. *Trends in Cognitive Sciences*, 22:294–306, 2018. doi: 10.1016/j.tics.2018.01.009.

[65] T. Parr and K. J. Friston. The anatomy of inference: Generative models and brain structure. *Frontiers in Computational Neuroscience*, 12:90, 2018. doi: 10.3389/fncom.2018.00090.

[66] M. Joffily and G. Coricelli. Emotional valence and the free-energy principle. *PLoS Computational Biology*, 9:e1003094, 2013. doi: 10.1371/journal.pcbi.1003094.

[67] C. Hesp, R. Smith, T. Parr, M. Allen, K. J. Friston, and M. J. D. Ramstead. Deeply felt affect: The emergence of valence in deep active inference. *Neural Computation*, 33: 398–446, 2021. doi: 10.1162/neco_a_01341.

[68] M. Miller, J. Kiverstein, and E. Rietveld. The predictive dynamics of happiness and well-being. *Emotion Review*, 14:15–30, 2022. doi: 10.1177/17540739211063851.

[69] A. K. Seth and K. J. Friston. Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B*, 371(1708):20160007, 2016. doi: 10.1098/rstb.2016.0007.

[70] V. Ainley, M. A. J. Apps, A. Fotopoulou, and M. Tsakiris. 'Bodily precision': A predictive coding account of individual differences in interoceptive accuracy. *Philosophical Transactions of the Royal Society B*, 371:20160003, 2016. doi: 10.1098/rstb.2016.0003.

[71] L. F. Barrett and W. K. Simmons. Interoceptive predictions in the brain. *Nature reviews neuroscience*, 16:419–429, 2015. doi: 10.1038/nrn3950.

[72] L. F. Barrett. The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, 12:1–23, 2017. doi: 10.1093/scan/nsw154.

[73] D. Parvizi-Wayne. How preference enslaves attention: calling into question the endogenous/exogenous distinction from an active inference perspective. *Phenomenology and the Cognitive Sciences*, 2024. doi: 10.1007/s11097-024-10028-5.

[74] C. Jicol, H. Y. Cheng, K. Petrini, and E. O'Neill. A predictive model for understanding the role of emotion for the formation of presence in virtual reality. *PLoS ONE*, 18: e0280390, 2023. doi: 10.1371/journal.pone.0280390.

[75] G. Deane, M. Miller, and S. Wilkinson. Losing ourselves: Active inference, depersonalization, and meditation. *Frontiers in Psychology*, 11:539726, 2020. doi: 10.3389/fpsyg.2020.539726.

[76] A. K. Seth. Presence, objecthood, and the phenomenology of predictive perception. *Cognitive Neuroscience*, 6:111–117, 2015. doi: 10.1080/17588928.2015.1026888.

[77] T. Metzinger. *Being No One: The Self-Model Theory of Subjectivity*. MIT Press, Cambridge, MA, 2003. doi: 10.7551/mitpress/1551.001.0001.

[78] J. Limanowski. Precision control for a flexible body representation. *Neuroscience & Biobehavioral Reviews*, 134:104401, 2022. doi: 10.1016/j.neubiorev.2021.10.023.

[79] G. Riva, B. K. Wiederhold, and F. Mantovani. Searching for the metaverse: Neuroscience of physical and digital communities. *Cyberpsychology, Behavior, and Social Networking*, 27:9–18, 2024. doi: 10.1089/cyber.2023.0040.

[80] M. Gonzalez-Franco and J. Lanier. Model of illusions and virtual reality. *Frontiers in Psychology*, 8:1125, 2017. doi: 10.3389/fpsyg.2017.01125.

[81] N. Dijkstra, P. Kok, and S. M. Fleming. Perceptual reality monitoring: Neural mechanisms dissociating imagination from reality. *Neuroscience & Biobehavioral Reviews*, 135:104557, 2022. doi: 10.1016/j.neubiorev.2022.104557.

[82] N. Dijkstra and S. M. Fleming. Subjective signal strength distinguishes reality from imagination. *Nature Communications*, 14:1627, 2023. doi: 10.1038/s41467-023-37322-1.

[83] M.-L. Ryan. Fiction, non-factuals, and the principle of minimal departure. *Poetics*, 9(4):403–422, 1980. ISSN 0304-422X. doi: 10.1016/0304-422X(80)90030-3. URL https://www.sciencedirect.com/science/article/pii/0304422X80900303.

[84] R. J. Gerrig and D. A. Prentice. The representation of fictional information. *Psychological Science*, 2(5):336–340, 1991. doi: 10.1111/j.1467-9280.1991.tb00162.x.

[85] M. Appel and T. Richter. Persuasive effects of fictional narratives increase over time. *Media Psychology*, 10(1):113–134, 2007. doi: 10.1080/15213260701301194.

[86] N. E. Seymour, A. G. Gallagher, S. A. Roman, M. K. O'brien, V. K. Bansal, D. K. Andersen, and R. M. Satava. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Annals of surgery*, 236(4):458–464, 2002. doi: 10.1097/01.SLA.0000028969.51489.B4.

[87] S. C. Michalski, A. Szpak, D. Saredakis, T. J. Ross, M. Billinghurst, and T. Loetscher. Getting your game on: Using virtual reality to improve real table tennis skills. *PloS one*, 14(9):e0222351, 2019. doi: 10.1371/journal.pone.0222351.

[88] D. Mobbs, C. Hagan, T. Dalgleish, B. Silston, and C. Prévost. The ecology of human fear: survival optimization and the nervous system. *Frontiers in Neuroscience*, 9:121062, 2015. doi: 10.3389/fnins.2015.00055.

[89] M. Olszanowski, N. Frankowska, and A. Tolopilo. "Rear bias" in spatial auditory perception: Attentional and affective vigilance to sounds occurring outside the visual field. *Psychophysiology*, 60:e14377, 2023. doi: 10.1111/psyp.14377.

[90] A. L. Francis and D. J. Strauss. An evolutionary history of human listening: Mismatches between auditory priors and contemporary acoustic environments. *JASA Express Letters*, 6(3):037202, 2026. doi: 10.1121/10.0042868.

[91] R. Desimone and J. Duncan. Neural mechanisms of selective visual attention. *Annual review of neuroscience*, 18:193–222, 1995. doi: 10.1146/annurev.ne.18.030195.001205.

[92] C. Trenado, L. Haab, and D. J. Strauss. Corticothalamic feedback dynamics for neural correlates of auditory selective attention. *IEEE Trans Neural Syst Rehabil Eng.*, 17:46–52, 2009. doi: 10.1109/TNSRE.2008.2010469.

[93] D. J. Strauss, F. I. Corona-Strauss, C. Trenado, C. Bernarding, W. Reith, M. Latzel, and M. Froehlich. Electrophysiological correlates of listening effort: Neurodynamical modeling and measurement. *Cogn. Neurodyn.*, 4, 2010. doi: 10.1007/s11571-010-9111-3.

[94] L. J. Speed, L. S. Eekhof, and M. Mak. The role of visual imagery in story reading: Evidence from aphantasia. *Consciousness and Cognition*, 118:103645, 2024. doi: 10.1016/j.concog.2024.103645.

[95] N. Abdelrahman, D. Melcher, and P. Ripollés. Seeing is feeling: How aphantasia alters emotional engagement with stories. *Psychophysiology*, 62(7):e70100, 2025. doi: 10.1111/psyp.70100.

[96] M. Mariam, W. Delb, F.I. Corona-Strauss, M. Bloching, and D.J. Strauss. Comparing the habituation of late auditory evoked potentials to loud and soft sound. *Physiol Meas*, 30:141–153, 2009. doi: 10.1088/0967-3334/30/2/003.

[97] L. Haab, C. Trenado, M. Mariam, and D. J. Strauss. Neurofunctional model of large-scale correlates of selective attention governed by stimulus-novelty. *Cogn. Neurodyn.*, 5 (1):103–111, 2011. doi: 10.1007/s11571-010-9150-9.

[98] Z. Mortezapouraghdam, L. Haab, F. I. Corona-Strauss, G. Steidl, and D. J. Strauss. Assessment of long-term habituation correlates in event-related potentials using a von Mises model. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 23(3):363–373, 2015. doi: 10.1109/TNSRE.2014.2361614.

[99] M. Sarter, W.J. Gehring, and R. Kozak. More attention must be paid: The neurobiology of attentional effort. *Brain Research Reviews*, 51:145–160, 2006. doi: 10.1016/j.brainresrev.2005.11.002.

[100] D. Kahneman. *Attention And Effort*. Prentice–Hall, Englewood Cliffs, NJ, 1973.

[101] T. Waltemate, I. Senna, F. Hülsmann, M. Rohde, S. Kopp, M. O. Ernst, and M. Botsch. The impact of latency on perceptual judgments and motor performance in closed-loop interaction in virtual reality. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology (VRST '16)*, pages 27–35, New York, NY, 2016. ACM. doi: 10.1145/2993369.2993381.

[102] D. J. Strauss, F. I. Corona-Strauss, A. Schroeer, P. Flotho, R. Hannemann, and S. A. Hackley. Vestigial auriculomotor activity indicates the direction of auditory attention in humans. *eLife*, 9:e54536, 2020. doi: 10.7554/eLife.54536.

[103] A. Schroeer, M. R. Andersen, M. L. Rank, R. Hannemann, E. B. Petersen, F. M. Rønne, D. J. Strauss, and F. I. Corona-Strauss. Assessment of vestigial auriculomotor activity to acoustic stimuli using electrodes in and around the ear. *Trends in Hearing*, 27: 23312165231200158, 2023. doi: 10.1177/23312165231200158.

[104] A. Schroeer, F. I. Corona-Strauss, R. Hannemann, S. A. Hackley, and D. J. Strauss. Electromyographic correlates of effortful listening in the vestigial auriculomotor system. *Frontiers in Neuroscience*, 18, 2024. doi: 10.3389/fnins.2024.1462507.

[105] L. Renz, R. A. Vargas, Z. Qiu, T. Grün, F. I. Corona-Strauss, A. J. Pegna, and D. J. Strauss. Fearful faces in peripheral vision: The vestigial auriculomotor system as an indicator of affective processing. *Brain Research Bulletin*, 241:111905, 2026. doi: 10.1016/j.brainresbull.2026.111905.

[106] S. Bonnaire, M. J. Bhamborae, D. J. Strauss, and F. I. Corona-Strauss. Unraveling affective states from the auriculomotor system. *IEEE Transactions on Affective Computing*, pages 1–12, 2026. doi: 10.1109/TAFFC.2026.3717435. early access.

[107] J.-P. Stauffert, F. Niebling, and M. E. Latoschik. Latency and cybersickness: Impact, causes, and measures. A review. *Frontiers in Virtual Reality*, 1:582204, 2020. doi: 10.3389/frvir.2020.582204.

[108] K. P. Körding, U. Beierholm, W. J. Ma, S. Quartz, J. B. Tenenbaum, and L. Shams. Causal inference in multisensory perception. *PLoS ONE*, 2:e943, 2007. doi: 10.1371/journal.pone.0000943.

[109] L. Shams and U. R. Beierholm. Causal inference in perception. *Trends in Cognitive Sciences*, 14:425–432, 2010. doi: 10.1016/j.tics.2010.07.001.

[110] M. O. Ernst and M. S. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, 2002. doi: 10.1038/415429a.

[111] J. T. Reason. Motion sickness adaptation: A neural mismatch model. *Journal of the Royal Society of Medicine*, 71:819–829, 1978. doi: 10.1177/014107687807101109.

[112] C. M. Oman. Motion sickness: A synthesis and evaluation of the sensory conflict theory. *Canadian Journal of Physiology and Pharmacology*, 68:294–303, 1990. doi: 10.1139/y90-044.

[113] S. Weech, S. Kenny, and M. Barnett-Cowan. Presence and cybersickness in virtual reality are negatively related: A review. *Frontiers in Psychology*, 10:158, 2019. doi: 10.3389/fpsyg.2019.00158.

[114] J. Limanowski and K. Friston. Active inference under visuo-proprioceptive conflict: Simulation and empirical results. *Scientific Reports*, 10:4010, 2020. doi: 10.1038/s41598-020-61097-w.

[115] H. Brown, R. A. Adams, I. Parees, M. Edwards, and K. Friston. Active inference, sensory attenuation and illusions. *Cognitive Processing*, 14:411–427, 2013. doi: 10.1007/s10339-013-0571-3.

[116] S. Palmisano, R. S. Allison, and J. Kim. Cybersickness in head-mounted displays is caused by differences in the user's virtual and physical head pose. *Frontiers in Virtual Reality*, 1:587698, 2020. doi: 10.3389/frvir.2020.587698.

[117] M. Botvinick and J. Cohen. Rubber hands 'feel' touch that eyes see. *Nature*, 391:756, 1998. doi: 10.1038/35784.

[118] M. Slater, B. Spanlang, M. V. Sanchez-Vives, and O. Blanke. First person experience of body transfer in virtual reality. *PLoS ONE*, 5:e10564, 2010. doi: 10.1371/journal.pone.0010564.

[119] D. Banakou, R. Groten, and M. Slater. Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences of the United States of America*, 110:12846–12851, 2013. doi: 10.1073/pnas.1306779110.

[120] A. Maselli and M. Slater. The building blocks of the full body ownership illusion. *Frontiers in Human Neuroscience*, 7:83, 2013. doi: 10.3389/fnhum.2013.00083.

[121] V. I. Petkova and H. H. Ehrsson. If I were you: Perceptual illusion of body swapping. *PLoS ONE*, 3:e3832, 2008. doi: 10.1371/journal.pone.0003832.

[122] B. Lenggenhager, T. Tadi, T. Metzinger, and O. Blanke. Video ergo sum: Manipulating bodily self-consciousness. *Science*, 317:1096–1099, 2007. doi: 10.1126/science.1143439.

[123] C. Spence and S. Squire. Multisensory integration: Maintaining the perception of synchrony. *Current Biology*, 13(13):R519–R521, 2003. doi: 10.1016/S0960-9822(03)00445-7.

[124] A. R. Powers, A. R. Hillock, and M. T. Wallace. Perceptual training narrows the temporal window of multisensory binding. *Journal of Neuroscience*, 29(39):12265–12274, 2009. doi: 10.1523/JNEUROSCI.3501-09.2009.

[125] P. Greenwood, M. Lin, R. Sundararajan, K. Fryxell, and R. Parasuraman. Synergistic effects of genetic variation in nicotinic and muscarinic receptors on visual attention but not working memory. *Proceedings of the National Academy of Sciences*, 106:3633–3638, 2009. doi: 10.1073/pnas.0807891106.

[126] A. P. Anokhin, S. Golosheykin, and A. C. Heath. Genetic and environmental influences on emotion-modulated startle reflex: A twin study. *Psychophysiology*, 44(1):106–112, 2007. doi: 10.1111/j.1469-8986.2006.00486.x.

[127] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators & Virtual Environments*, 7(3):225 – 240, 1998. doi: 10.1162/105474698565686.

[128] J. M. Rigby, D. P. Brumby, S. J. J. Gould, and A. L. Cox. Development of a questionnaire to measure immersion in video media: The Film IEQ. In *Proceedings of the 2019 ACM international conference on interactive experiences for TV and online video*, pages 35–46, 2019. doi: 10.1145/3317697.3323361.

[129] C. Jennett, A. L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, and A. Walton. Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, 66(9):641–661, 2008. ISSN 1071-5819. doi: 10.1016/j.ijhcs.2008.04.004.

[130] R. Näätänen, P. Paavilainen, T. Rinne, and K. Alho. The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, 118:2544–2590, 2007. doi: 10.1016/j.clinph.2007.04.026.

[131] J. Polich. Updating P300: an integrative theory of P3a and P3b. *Clin Neurophysiol.*, 118:2128–2148, 2007. doi: 10.1016/j.clinph.2007.04.019.

[132] S. A. Hillyard, E. K. Vogel, and S. J. Luck. Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 353:1257–1270, 1998. doi: 10.1098/rstb.1998.0281.

[133] A. Mai, S. A. Hillyard, D. J. Strauss, and F. I. Corona-Strauss. Selective listening to unpredictable sound sequences increases tonic muscle activity in the human vestigial auriculomotor system. *eNeuro*, 12(10):ENEURO.0301–25.2025, 2025. doi: 10.1523/ENEURO.0301-25.2025.

[134] A. Schroeer, F. I. Corona-Strauss, R. Hannemann, S. A. Hackley, and D. J. Strauss. The vestigial pinna-orienting system in humans briefly suppresses superior auricular muscle activity during reflexive orienting towards auditory stimuli. *Journal of Neurophysiology*, 132:514–526, 2024. doi: 10.1152/jn.00024.2024.

[135] A. Van Boxtel. Facial EMG as a tool for inferring affective states. In *Proceedings of Measuring Behavior 2010*, pages 104–108, Eindhoven, The Netherlands, 2010.

[136] S. Bonnaire, F. I. Corona-Strauss, and D. J. Strauss. The rise of super-emotional intelligence. *Frontiers in Artificial Intelligence*, 9:1975692, 2026. doi: 10.3389/frai.2026.1975692.

[137] E. Siegel, M. K. Sands, W. Van den Noortgate, P. Condon, Y. Chang, J. Dy, K. S. Quigley, and L. F. Barrett. Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories. *Psychological Bulletin*, 144(4): 343–393, 2018. doi: 10.1037/bul0000128.

[138] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak. Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, 20(1):1–68, 2019. doi: 10.1177/1529100619832930.

[139] E. N. Schneider, B. Buchheit, P. Flotho, M. J. Bhamborae, F. I. Corona-Strauss, F. Dauth, M. Alayan, and D. J. Strauss. Electrodermal responses to driving maneuvers in a motion sickness inducing real-world driving scenario. *IEEE Transactions on Human-Machine Systems*, 52:994–1003, 2022. doi: 10.1109/THMS.2022.3188924.

[140] N. Bouizegarene, M. J. D. Ramstead, A. Constant, K. J. Friston, and L. J. Kirmayer. Narrative as active inference: An integrative account of cognitive and social functions in adaptation. *Frontiers in Psychology*, 15:1345480, 2024. doi: 10.3389/fpsyg.2024.1345480.

[141] Y. Yeshurun, S. Swanson, E. Simony, J. Chen, C. Lazaridi, C. J. Honey, and U. Hasson. Same story, different story: the neural representation of interpretive frameworks. *Psychological science*, 28(3):307–319, 2017.

[142] E. S. Finn, P. R. Corlett, G. Chen, P. A. Bandettini, and R. T. Constable. Trait paranoia shapes inter-subject synchrony in brain activity during an ambiguous social narrative. *Nature communications*, 9(1):2043, 2018.

[143] G. Braun, Y. Yeshurun, and E. Shetreet. Shared disbelief and shared belief: Belief and disbelief as drivers of interpersonal neural synchronization during narrative processing. *Proceedings of the National Academy of Sciences*, 122(23):e2422396122, 2025. doi: 10.1177/0956797616682029.

[144] M. Nguyen, T. Vanderwal, and U. Hasson. Shared understanding of narratives is correlated with shared neural responses. *NeuroImage*, 184:161–170, 2019. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2018.09.010.

[145] A. Tellegen and G. Atkinson. Openness to absorbing and self-altering experiences (" absorption"), a trait related to hypnotic susceptibility. *Journal of abnormal psychology*, 83(3):268–277, 1974. doi: 10.1037/h0036681.

[146] G. A. Calvert, C. Spence, and B. E. Stein, editors. *The handbook of multisensory processes*. MIT Press, 2004. doi: 10.7551/mitpress/3422.001.0001.

[147] C. Spence and V. Santangelo. Capturing spatial attention with multisensory cues: a review. *Hear Res.*, 258:134–142, 2009. doi: 10.1016/j.heares.2009.04.015.

[148] D. Thinnes, J. Trapero, E. Wagner, and D. J. Strauss. Neuroergonomics in digital operating rooms: Applying the two-competitor model of attention to the surgical context. *IEEE Transactions on Human-Machine Systems*, 55(5):765–776, October 2025. doi: 10.1109/THMS.2025.3601222.

[149] T. A. Chowdhury, A. L. Francis, and D. J. Strauss. Using a surrogate human digital twin to model embodied perception and cognition for neuroergonomics. *Frontiers in Neuroergonomics*, 7:1939412, 2026. doi: 10.3389/fnrgo.2026.1939412.

[150] C. J. Bohil, B. Alicea, and F. A. Biocca. Virtual reality in neuroscience research and therapy. *Nature Reviews Neuroscience*, 12:752–762, 2011. doi: 10.1038/nrn3122.

[151] N. Sajid, P. Ball, T. Parr, and K. Friston. Active inference: Demystified and compared. *Neural Computation*, 33(3):674–712, 2021. doi: 10.1162/neco_a_01357.

[152] K. J. Friston, T. Parr, and B. de Vries. The graphical brain: Belief propagation and active inference. *Network Neuroscience*, 1:381–414, 2017. doi: 10.1162/NETN_a_00018.

[153] M. B. Mirza, R. A. Adams, C. D. Mathys, and K. J. Friston. Scene construction, visual foraging, and active inference. *Frontiers in Computational Neuroscience*, 10:56, 2016. doi: 10.3389/fncom.2016.00056.