philosophies

MDPI

Article

Facing Immersive "Post-Truth" in AIVR?

Nadisha-Marie Aliman [1,*] and Leon Kester [2]

[1] Department of Information and Computing Sciences, Utrecht University, 3584 CC Utrecht, The Netherlands
[2] TNO Netherlands, 2597 AK The Hague, The Netherlands; leon.kester@tno.nl
* Correspondence: nadishamarie.aliman@gmail.com

Received: 3 November 2020 ; Accepted: 11 December 2020; Published: 15 December 2020

Abstract: In recent years, prevalent global societal issues related to fake news, fakery, misinformation, and disinformation were brought to the fore, leading to the construction of descriptive labels such as "post-truth" to refer to the supposedly new emerging era. Thereby, the (mis-)use of technologies such as AI and VR has been argued to potentially fuel this new loss of "ground-truth", for instance, via the ethically relevant deepfakes phenomena and the creation of realistic fake worlds, presumably undermining experiential veracity. Indeed, *unethical* and malicious actors could harness tools at the intersection of AI and VR (AIVR) to craft what we call *immersive falsehood*, fake immersive reality landscapes deliberately constructed for malicious ends. This short paper analyzes the ethically relevant nature of the background against which such malicious designs in AIVR could exacerbate the intentional proliferation of deceptions and falsities. We offer a reappraisal expounding that while immersive falsehood could manipulate and severely jeopardize the inherently affective constructions of social reality and considerably complicate falsification processes, humans may neither inhabit a post-truth nor a post-falsification age. Finally, we provide incentives for future AIVR safety work, ideally contributing to a future era of *technology-augmented critical thinking*.

Keywords: AI; VR; AI safety; AIVR safety; disinformation; affective science; psychology; scientific method; epistemology

1. Motivation

In the last few years, the information ecosystem was permeated by falsehood-related concepts, such as fake news [1], deepfakes [2], fake realities [3], and digital fakery [4], as well as more globally fake science [5] and *post-truths* [6]. Regarding fakery and truth in extended reality (XR) settings and thus by extension VR, Slater et al. [7] recently argued that: *"Society is based on the premise that sensory experiences give ground truth. XR at societal scales has the capacity to decouple sensory experience from ground truth, potentially undermining some core elements of social fabric"*. Moreover, it has been stated that the deployment of AI deepfakes may foster the acquisition of false memories [8]. This could be conceivably exacerbated within future extensions of technically already feasible "VR deepfakes" [9–11] by the particular aptness of VR to facilitate durable memories [12]. While such issues would already play a role regarding unintentional failure modes elicited by ethically aware actors in AIVR, recent research related to the security and safety of AI [13–16] and VR [17–20] respectively emphasizes the need to additionally consider the presence of *unethical* malicious actors. Thereby, to consider intentional malevolent design in AIVR could offer a worst-case scenario analysis [21] that can shed more light on the extent of potential consequences exhibited by the deployment of AIVR technology, but also by simpler cases in AI and VR separately. For instance, when addressing defense methods against AI-generated fakery in future immersive (VR) journalism contexts for disinformation purposes [22], one might gain insights on how to tackle the exposition to non-immersive deepfake artefacts. Simultaneously, it might help to foster the vulnerability awareness of VR users yielding cautionary attitudes towards manipulation.

In this paper, we focus on immersive falsehood in AIVR [22], the deliberate construction of fake immersive reality landscapes for malicious ends. Using this example, we contemplate the following question: *"Can malicious actors in AIVR exacerbate the presumed post-truth phenomenon via immersive falsehood?"*. Decisively, our answer to this question is that it is the wrong question to ask, for various reasons that require to be elucidated. While throughout history, many "rational" traditions were averse to affective motives and attempted to distance themselves from visceral and bodily elements, modern affective science assumes that affect is an inseparable part of cognition and perception [23,24]. Moreover, VR settings are known for their profound affective impacts on users [25]. Hence, Section 2 first elaborates on the epistemological implications of affect as an intrinsic ingredient in human cognition and perception—not only in VR. Extending beyond that, Section 3 explains why the term "post-truth" may *not* serve as an accurate description of the current age. Moreover, VR has been described to offer a rich counterfactual experiential testbed for ethics in technological contexts, such as AI [26,27]. Building on this, Section 4 briefly discusses how future affective computing and virtual reality methods could be harnessed for counterfactual and other measures that seek to allow an understanding and debiasing of one's own constructions as a response to immersive falsehood. Finally, Section 5 concludes.

2. Nested Affective VR Worlds

Before addressing the previously mentioned question related to AIVR and "post-truth", it might be essential to first collate information on the nature of the human perception of reality from a transdisciplinary perspective. As famously stated by Feynman [28]: *"Science is a way of trying not to fool yourself"*. Against the background of post-truth claims, it seems important to first carefully deconstruct the notion of "ground truth" in different, ethically relevant human contexts:

- **Affective Realism and Social Reality:** As stated by Barrett, *"The human brain is anatomically structured so that no decision or action can be free of interoception and affect"* [29]. Thereby, interoception and interoceptive predictions pertain to statistical regularities of the internal milieu of an organism (related to the *body*) [30], while core affect is seen as a fundamental property of consciousness [31] with especially valence (pleasant/unpleasant) and arousal (activated/deactivated) as crucial components. To put it very simply, according to constructionist theories in psychology, all mental states are based on constructions involving three basic elements [32]: exteroceptive sensory array (related to sensory predictions and information sampled from external world), interoception, and prior knowledge including past experience. The hereto linked circumstance that *"affective feelings (incidental or not) naturally infuse our perceptions and give us a sense of confidence that they are valid windows onto the real world"* [33] has been termed *affective realism*. Thereby, human perception imposes cognitive-affective concepts on the world, often previously constructed in social reality (abbreviated with SR in the following) and shared via language. In this sense, human perception also exhibits a biologically shaped social nature, given that humans reciprocally regulate the biological nervous systems of their social conspecifics[1] [37,39,40,42] via interpersonal physiological dynamics [35] that humans can even remotely bring about using language [29]. Generally, *"human brains are transactive and cannot be considered outside the context of other human brains"* [37]. In our view, affective realism and the embodied nature of cognition are crucial to further understanding, as it stresses that SR is of an embodied and *perceiver-dependent* nature [43]—as are mental constructions like emotions [31], moral judgments [44], thoughts, perceptions, and so on.

[1] For instance, social groups reveal an attunement of physiological parameters [34,35], social relationships act as physiological regulators [34,36], biobehavorial synchrony serves as scaffold for the maturation of infant brains facilitating social development [37,38] and the metabolic costs and benefits of interpersonal physiological dynamics modulate social interactions throughout a lifetime [39,40]. Hence, it is also no surprise that social isolation comes with the physiological burden of less regulatory facilitations [29] and *"lacking social connection qualifies as a risk factor for premature mortality"* [41].

- **Theory-Ladenness:** In science, it is important to separate perceiver-dependent from *perceiver-independent* phenomena which directly pertain to the physical reality (abbreviated with PhyR in the following) that diverse scientific areas attempt to understand. As emphasized by Barrett, *"all science relies on human concepts, and this is true for the astronomy as it is for the science of emotion"* [45]. For illustrative purposes, Barrett explains that while the existence of celestial bodies in PhyR is perceiver-independent, the status of one celestial body being a planet is not (see the reclassification of Pluto from planet to dwarf planet). In short, humans do not have direct access to the hidden states in PhyR, but try to infer those. In this process, one needs to keep in mind that all observations are *theory-laden*, which cautions scientists that since one actively samples the environment to gather data, one's prior socio-cultural context, hypotheses, and affective predispositions inherently shape what we perceive as information and what as noise. To conclude, even prior to AIVR, our perception of reality was never entirely objective, nor did we directly have access to truth which could suddenly get lost by experiencing immersive falsehood. SR is as real as socio-cultural conventions, such as language or money. While its embodied constructions contain real physiological ingredients grounded in PhyR, one often tends to see more of what one believes than vice versa [46,47].
- **Nested VR Ground Truth:** An important phenomenological aspect of human experience is its virtual, perspectival, and egocentric nature [48,49] with a simultaneous grounding in PhyR linked to cybernetic control [50]. It has been postulated that the human persona inhabits a virtual world generated by the brain [48,50] and governed by affective dynamics to navigate the physical environment anticipating bodily needs before they occur (this process has also been termed allostasis [30,51]). More generally, waking time, imagination, and dreaming are all assumed to be linked to a virtual reality experience (that we abbreviate as VR_{Mind} in the following) which is generated by the brain for embodied control purposes [50]. In waking time, this virtual experience is directly constrained by PhyR, while dreaming has been described as *"virtual reality proper"* [48] due to the decoupling from external sensory stimulation and blockage of motor actuators (with the exception of e.g., eye muscles). While awake and wearing technical VR headsets and being immersed in a virtual world, a complex novel nested situation occurs, *"a nested form of information flow in which the biological mind and its technological niche influence each other in ways we are just beginning to understand"* [52]. In these scenarios, our VR_{Mind} experience is constrained by both the artificially created VR world, and still partially always also PhyR (e.g., simply by having a body and literally sitting, standing or walking during the setting). In short, even without using any VR technology, the human experience of the world does not only reflect statistical regularities about PhyR, but consists in goal-directed, embodied, affective, and theory-laden virtual constructions of a perspectival and perceiver-dependent nature, such as those involved in SR. When using VR, one adds an additional layer of sensory-motor and affective constraints, leading to a nested composition. With social VR [53], a novel special case of SR constrained by VR arises and poses new challenges.

3. Immersive Falsehood–Post-Truth, Post-Falsification or Other?

After having analyzed various relevant aspects related to the human perception of reality, one can now re-examine the initial question on whether malicious actors in AIVR could exacerbate assumed "post-truth" phenomena via immersive falsehood.

- ***Post-Truth?*** As advanced by Buffachi [6], the perception of a "post-truth" era may be linked to the definition assigned to truth in the first place—especially when truth is associated with *consensus*, which seems to be compromised in modern times. We agree with Buffachi to instead utilize the word "truth" in a much more deflationary manner, namely, strictly for scientific endeavors. In our view, consensus is a dominant factor in SR, and technologies such as AIVR may be able to profoundly distort features of SR and certain democratic processes. However, when it comes to PhyR, it is obvious that AIVR artefacts do not irreversibly destroy our capability to create refutable

conjectures about PhyR. While one could believe that the loss of truth would be exacerbated by AIVR because observations may become unreliable[2], it is important to keep in mind that no repetition of observations can ever provide an experimental logically valid justification for a theory [55,56]. As Karl Popper explained, *induction is logically invalid*, and for instance, no amount of observed white swans ever proves that all swans are white [55,58]. He pointed at the asymmetry between falsifiability and verifiability [59], emphasizing falsifiability as one of the most important criteria for scientific theories. While no amount of successful experiments can ever justify a theory, that is, establish its truth, negative experiments can make the theory problematic. (Thereby, note that as elaborated in the Duhem–Quine thesis [60], no experimental falsification attempt can be considered as absolute and conclusive. Consequently, it is the case in practice that only multiple contextualized failures and/or the presence of competitive alternatives contribute to consider the theory as refuted. However, since justifications are logically invalid *on principle* [55,56,59], this type of more complex, context-aware, sophisticated falsificationalism and criticism remains the recommendable alternative.) In short, if one does neither equate truth with social consensus nor scientific truth with justification via observations, immersive falsehood of the future lets the existence of truth be untouched—even if not directly accessible. Hence, *there is no reason to assume that humans inhabit a post-truth era*. However, this very asymmetry between falsifiability and verifiability leads to a further complication addressed in the next point.

- ***Post-Falsification?*** In our view, a legitimate concern is the ability of malevolent actors in AIVR to compromise material that could be utilized to falsify hypotheses in diverse contexts, such as science, history, forensics, and journalism with political repercussions. As stated by Popper, while coherence cannot attest truth, *"inconsistency and incoherence do establish falsehood"* [55]. Concerning historical and also forensic sources [61], it is important to analyze whether they exhibit mutual or internal inconsistencies. In other scientific areas, falsification attempts can be more easily repeated, but scientists often rely at least on the honesty of other entities publishing their experimental results (i.e., that other scientists do not deliberately temper their results). For instance, future immersive falsehood in the form of AI-manipulated VR news for disinformation [22], but also defamation and extortion purposes, could distort historical and forensic records and exacerbate issues in the information ecosystem. Malicious actors could craft future realistic immersive experiences (e.g., of fake AI-generated confirmatory experiments and research [57]) to undermine the scientific enterprise. With increasing degrees of realism, many scientists may not stay immune against such strategies. At first sight, it might thus seem as if immersive falsehood could compromise falsification (e.g., via future VR deepfakes [9,10]). Fake memories could be specifically induced in users [22] that may turn out to be difficult to detect. However, as noted under the last bullet point and known from the Duhem–Quine thesis, it is *not* the case that falsification can be experimentally conclusively established *in isolation* (mainly due to inherent background assumptions that always play a role). In this vein, it signifies that immersive falsehood would predominantly complicate the falsification process by having the potential to lure humans into wrong background assumptions and slowing down progress. However, while acknowledging these significant impacts of immersive falsehood, this complication seems, however, to represent *a matter of degree*, rather than a matter of kind, which is why we postulate that *there is no reason to assume the science-threatening scenario of a post-falsification era*.

---

[2] In fact, from a Bayesian empiricist point of view which links science to *true beliefs* and *empirical justifications*, deepfakes are already assumed to represent *epistemic threats* [54] gradually emptying audiovisual samples of information. By contrast, Popperian epistemology [55] sees science as an *explanation-based* and *criticism-centered* endeavor with *falsifiability* as decisive criterium—which has been extended by Deutsch [56] who views science as the quest to identify invariant *hard-to-vary explanations* of reality. On that view, deepfakes (and immersive falsehood) do *not* put truth at risk (see [57] for more details including the safety-relevant urgency to thematize these fundamental Bayesian vs. Popperian epistemic divergences).

4. Future Work

In the light of this complex and nuanced landscape related to the worst-case consequences of immersive falsehood, future work could address transdisciplinary countermeasures. For instance, while legal and technical strategies could *proactively* attempt to penalize, detect, and establish accountability for immersive falsehood artefacts which might stay a controversial issue, one might also need to anticipate the unavoidable proliferation of at least a part of those within the complex, heterogeneous, and dense information ecosystem. Therefore, it may be of importance to additionally develop *reactive* strategies addressing the issue on how individuals can retrospectively counter having *already experienced* samples of immersive falsehood without their knowledge [22]. First, one may, for instance, need to consciously entertain stronger doubts towards visceral and affective experiences. For this purpose, real-time affective monitoring [62,63], such as during the consumption of immersive journalism and VR news, could be investigated. By way of example, measurable physiological arousal parameters [64–66] could be visually displayed to the user to encourage a critical stance towards the experienced contents. Second, a type of counterfactual awareness training in VR may promote critical scrutiny by exposing users to design fiction scenarios featuring a conjunction of immersive news samples related to *real* events on the one hand, and fake ones based on *plausible counterfactuals* on the other [22]. Third, users could experience immersive counterfactual scenarios, illustrating the consequences of triggered doubts through AI-generated fakery in immersive or non-immersive news settings, and the dangers of false memory uptake. In fact, the *mere existence of deepfakes* has already led to doubts with lethal risk potentials, such as in the context of a failed military coup amidst pre-existing political unrest in Gabon [67,68]. By making the vulnerability of humans to these sorts of doubts and false memory constructions more palpable, user vigilance might consequently increase in AIVR contexts. This could also be supported via tailored VR experiences successfully elicitating *mortality salience* [69] (i.e., the awareness of one's mortality), which can motivate *safer* attitudes and behaviors [69–71]. VR could thus represent a suitable awareness-raising tool for future severe AI(VR) safety risks, such as by facilitating valuable retrospective counterfactual analyses [72]. Fourth, a generic recommendation that may already be applicable nowadays is to deliberately turn the confirmation bias [73] automatically reinforced via AI-empowered social media [74] *against itself* [57]. For example, one could create social media spaces (subsuming future social VR) that reinforce *critical thinking, life-long learning, and criticism* [57], which could be *deliberately* fueled via artificial bots (or non-player characters in VR), steering attention towards those patterns. Even if immersive falsehood would often not be resolved quickly, (AI-aided) social peer pressure reinforcing critical thinking and a focus on invariant good *explanations* could represent a necessarily incomplete, but principled defense.

5. Conclusions

In this short paper, we analyzed the extent to which malicious actors in AIVR could compromise truth across diverse areas, from societal contexts to science. In the light of affective realism and the perceiver-dependent nature of social reality, we deconstructed the nature of the term "ground truth" often prematurely assigned to the human experiential world. In a nutshell, we concluded that on a more strict deflationary account of truth linked to science and *not consensus*, we do *not* inhabit a post-truth era. First, humans were never equipped with direct access to physical reality in the first place. Second, the goal in science should, in any case, not consist in attempting to empirically identify and justify truth, because neither positive evidence nor consensus ever establishes truth, as put forth by Popper. Instead, the scientific method ideally heavily relies on *falsifiability* and error correction. In a further step, we thus analyzed whether falsifiability could be irreversibly endangered by immersive falsehood. Our analysis suggests that while the speed of falsification procedures could be considerably slowed down (which could generate serious complications in a broad range of domains including science, law enforcement, journalism, and politics), it would be *a matter of degree* and not of kind. Generally, whatever level of deception and disinformation is achieved by malicious actors, it does *not* per se eradicate the scientific method, and we likewise do

not inhabit a post-falsification era. A general epistemic view on science compatible with this is to conceive of it as an endless error-corrected quest for invariant hard-to-vary theoretical explanations of reality, as advocated by Deutsch [56]—a quest which can obviously *not* be terminally disrupted by slowed-down experimental falsification procedures. Last but not least, we proposed to defend against and face immersive falsehood by utilizing AIVR safety tools offering a rich counterfactual experiential testbed [22,26,75]. Ideally, these methods could contribute to what one could call a renewed counterfactual era of *technology-augmented critical thinking*. In short, while immersive falsehood neither terminally disrupts truth nor falsification, technology-augmented critical thinking (and concurrently, a dynamic augmentation of creativity [76] to craft novel unpredictable requisite solutions) seems indispensable in the light of various remaining severe socio-psycho-technological risks [22] that future immersive falsehood could involve and reinforce. Conceivable future risk examples could range from AI- [77] and VR-enabled [10] crimes to false memory constructions [7,8,22] over political unrest and safety-critical polarization in social media [78] (subsuming future social VR).

Author Contributions: N.-M.A. developed the main concepts of the paper and wrote the original draft. L.K. contributed via critical reflections and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lazer, D.M.; Baum, M.A.; Benkler, Y.; Berinsky, A.J.; Greenhill, K.M.; Menczer, F.; Metzger, M.J.; Nyhan, B.; Pennycook, G.; Rothschild, D.; et al. The science of fake news. *Science* **2018**, *359*, 1094–1096. [CrossRef] [PubMed]
2. Schick, N. *Deep Fakes and the Infocalypse: What You Urgently Need To Know*; Octopus: London, UK, 2020.
3. Polya, G. Fake news: "Fake realities" and lying by omission. *Glob. Res.* **2018**, *18*, 1.
4. Farid, H. Digital forensics in a post-truth age. *Comment. Forensic Sci. Int.* **2018**, *289*, 268–269. [CrossRef] [PubMed]
5. Hopf, H.; Krief, A.; Mehta, G.; Matlin, S.A. Fake science and the knowledge crisis: ignorance can be fatal. *R. Soc. Open Sci.* **2019**, *6*, 190161. [CrossRef]
6. Bufacchi, V. Truth, lies and tweets: A consensus theory of post-truth. *Philos. Soc. Crit.* **2020**. [CrossRef]
7. Slater, M.; Gonzalez-Liencres, C.; Haggard, P.; Vinkers, C.; Gregory-Clarke, R.; Jelley, S.; Watson, Z.; Breen, G.; Schwarz, R.; Steptoe, W.; et al. The ethics of realism in virtual and augmented reality. *Front. Virtual Real.* **2020**. [CrossRef]
8. Liv, N.; Greenbaum, D. Deep Fakes and Memory Malleability: False Memories in the Service of Fake News. *AJOB Neurosci.* **2020**, *11*, 96–104. [CrossRef]
9. Bose, A.J.; Aarabi, P. Virtual Fakes: DeepFakes for Virtual Reality. In Proceedings of the 2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP), Kuala Lumpur, Malaysia, 27–29 September 2019; p. 1.
10. Cole, S.; Maiberg, E. Deepfake Porn Is Evolving to Give People Total Control Over Women's Bodies. Available online: https://www.vice.com/en_uk/article/9keen8/deepfake-porn-is-evolving-to-give-people-total-control-over-womens-bodies (accessed on 4 August 2020).
11. Macaulay, T. New AR App Will Let You Model a Virtual Companion on Anyone You Want. Available online: https://thenextweb.com/neural/2020/06/01/new-ar-app-will-let-you-model-a-virtual-companion-on-anyone-you-want/ (accessed on 4 August 2020).
12. Krokos, E.; Plaisant, C.; Varshney, A. Virtual memory palaces: immersion aids recall. *Virtual Real.* **2019**, *23*, 1–15. [CrossRef]
13. Aliman, N.M.; Elands, P.; Hürst, W.; Kester, L.; Thórisson, K.R.; Werkhoven, P.; Yampolskiy, R.; Ziesche, S. Error-Correction for AI Safety. In *International Conference on Artificial General Intelligence*; Springer: Cham, Switzerland, 2020; pp. 12–22.
14. Brundage, M.; Avin, S.; Clark, J.; Toner, H.; Eckersley, P.; Garfinkel, B.; Dafoe, A.; Scharre, P.; Zeitzoff, T.; Filar, B.; et al. *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*; Future of Humanity Institute: Oxford, UK, 2018.

15. Pistono, F.; Yampolskiy, R.V. *Unethical Research: How to Create a Malevolent Artificial Intelligence*; Vernon Press: Wilmington, DE, USA, 2016; pp. 1–7.
16. Yampolskiy, R.V.; Spellchecker, M. Artificial intelligence safety and cybersecurity: A timeline of AI failures. *arXiv* **2016**, arXiv:1610.07997.
17. Pearlman, K. *Virtual Reality Brings Real Risks: Are We Ready?* USENIX Association: Berkeley, CA, USA, 2020.
18. Casey, P.; Baggili, I.; Yarramreddy, A. Immersive virtual reality attacks and the human joystick. *IEEE Trans. Dependable Secur. Comput.* **2019**. [CrossRef]
19. Gulhane, A.; Vyas, A.; Mitra, R.; Oruche, R.; Hoefer, G.; Valluripally, S.; Calyam, P.; Hoque, K.A. Security, Privacy and Safety Risk Assessment for Virtual Reality Learning Environment Applications. In Proceedings of the 2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 11–14 January 2019; pp. 1–9.
20. UW Allen School Security and Privacy Research Lab. 2019 Industry-Academia Summit on Mixed Reality Security, Privacy, and Safety: Summit Report. Available online: https://ar-sec.cs.washington.edu/research.html (accessed on 4 August 2020).
21. Happa, J.; Glencross, M.; Steed, A. Cyber security threats and challenges in collaborative mixed-reality. *Front. ICT* **2019**, *6*, 5. [CrossRef]
22. Aliman, N.M.; Kester, L. Malicious Design in AIVR, Falsehood and Cybersecurity-oriented Immersive Defenses. In Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), Utrecht, The Netherlands, 14–18 December 2020.
23. Barrett, L.F.; Simmons, W.K. Interoceptive predictions in the brain. *Nat. Rev. Neurosci.* **2015**, *16*, 419. [CrossRef] [PubMed]
24. Hoemann, K.; Feldman Barrett, L. Concepts dissolve artificial boundaries in the study of emotion and cognition, uniting body, brain, and mind. *Cogn. Emot.* **2019**, *33*, 67–76. [CrossRef] [PubMed]
25. Bujić, M.; Salminen, M.; Macey, J.; Hamari, J. "Empathy machine": How virtual reality affects human rights attitudes. *Internet Res.* **2020**, *30*, 1407–1425. [CrossRef]
26. Aliman, N.M.; Kester, L. Extending Socio-Technological Reality for Ethics in Artificial Intelligent Systems. In Proceedings of the 2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), San Diego, CA, USA, 9–11 December 2019; pp. 275–2757.
27. Aliman, N.M.; Kester, L.; Werkhoven, P. XR for Augmented Utilitarianism. In Proceedings of the 2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), San Diego, CA, USA, 9–11 December 2019; pp. 283–2832.
28. Feynman, R. Cargo cult science: Some remarks on science, pseudoscience, and learning how to not fool yourself-the 1974 Caltech commencement address. In *The Pleasure of Finding Things Out: The Best Short Works of Richard P. Feynman*; Basic Books: New York, NY, USA, 2005; pp. 205–216.
29. Barrett, L.F. *How Emotions are Made: The Secret Life of the Brain*; Houghton Mifflin Harcourt: Boston, MA, USA, 2017.
30. Kleckner, I.R.; Zhang, J.; Touroutoglou, A.; Chanes, L.; Xia, C.; Simmons, W.K.; Quigley, K.S.; Dickerson, B.C.; Barrett, L.F. Evidence for a large-scale brain system supporting allostasis and interoception in humans. *Nat. Hum. Behav.* **2017**, *1*, 1–14. [CrossRef]
31. Barrett, L.F. The theory of constructed emotion: an active inference account of interoception and categorization. *Soc. Cogn. Affect. Neurosci.* **2017**, *12*, 1–23. [CrossRef]
32. Oosterwijk, S.; Lindquist, K.A.; Anderson, E.; Dautoff, R.; Moriguchi, Y.; Barrett, L.F. States of mind: Emotions, body feelings, and thoughts share distributed neural networks. *NeuroImage* **2012**, *62*, 2110–2128. [CrossRef]
33. Wormwood, J.B.; Siegel, E.H.; Kopec, J.; Quigley, K.S.; Barrett, L.F. You are what I feel: A test of the affective realism hypothesis. *Emotion* **2019**, *19*, 788. [CrossRef]
34. Saxbe, D.E.; Beckes, L.; Stoycos, S.A.; Coan, J.A. Social Allostasis and Social Allostatic Load: A New Model for Research in Social Dynamics, Stress, and Health. *Perspect. Psychol. Sci.* **2020**, *15*, 469–482. [CrossRef]
35. Palumbo, R.V.; Marraccini, M.E.; Weyandt, L.L.; Wilder-Smith, O.; McGee, H.A.; Liu, S.; Goodwin, M.S. Interpersonal autonomic physiology: A systematic review of the literature. *Personal. Soc. Psychol. Rev.* **2017**, *21*, 99–141. [CrossRef]
36. Field, T. Relationships as regulators. *Psychology* **2012**, *3*, 467. [CrossRef]

37. Atzil, S.; Gao, W.; Fradkin, I.; Barrett, L.F. Growing a social brain. *Nat. Hum. Behav.* **2018**, *2*, 624–636. [CrossRef] [PubMed]
38. Feldman, R. Bio-behavioral synchrony: A model for integrating biological and microsocial behavioral processes in the study of parenting. *Parenting* **2012**, *12*, 154–164. [CrossRef]
39. Barrett, L.F.; Quigley, K.S.; Hamilton, P. An active inference theory of allostasis and interoception in depression. *Philos. Trans. R. Soc. B Biol. Sci.* **2016**, *371*, 20160011. [CrossRef]
40. Theriault, J.E.; Young, L.; Barrett, L.F. The sense of should: A biologically-based framework for modeling social pressure. *Phys. Life Rev.* **2020**, in press. [CrossRef]
41. Holt-Lunstad, J. Why social relationships are important for physical health: A systems approach to understanding and modifying risk and protection. *Annu. Rev. Psychol.* **2018**, *69*, 437–458. [CrossRef]
42. Sbarra, D.A.; Hazan, C. Coregulation, dysregulation, self-regulation: An integrative analysis and empirical agenda for understanding adult attachment, separation, loss, and recovery. *Personal. Soc. Psychol. Rev.* **2008**, *12*, 141–167. [CrossRef]
43. Barrett, L.F. Emotions are real. *Emotion* **2012**, *12*, 413. [CrossRef]
44. Gray, K.; Schein, C.; Cameron, C.D. How to think about emotion and morality: Circles, not arrows. *Curr. Opin. Psychol.* **2017**, *17*, 41–46. [CrossRef]
45. Barrett, L.F. Functionalism cannot save the classical view of emotion. *Soc. Cogn. Affect. Neurosci.* **2017**, *12*, 34–36. [CrossRef]
46. Fridman, J.; Barrett, L.F.; Wormwood, J.B.; Quigley, K.S. Applying the theory of constructed emotion to police decision making. *Front. Psychol.* **2019**, *10*, 1946. [CrossRef] [PubMed]
47. Kruglanski, A.W.; Jasko, K.; Friston, K. All Thinking is 'Wishful' Thinking. *Trends Cogn. Sci.* **2020**, *24*, 413–424. [CrossRef] [PubMed]
48. Hobson, J.A.; Hong, C.C.H.; Friston, K.J. Virtual reality and consciousness inference in dreaming. *Front. Psychol.* **2014**, *5*, 1133. [CrossRef] [PubMed]
49. Rudrauf, D.; Bennequin, D.; Williford, K. The Moon illusion explained by the projective consciousness model. *J. Theor. Biol.* **2020**, *507*, 110455. [CrossRef] [PubMed]
50. Williford, K.; Bennequin, D.; Friston, K.; Rudrauf, D. The projective consciousness model and phenomenal selfhood. *Front. Psychol.* **2018**, *9*, 2571. [CrossRef] [PubMed]
51. Corcoran, A.W.; Pezzulo, G.; Hohwy, J. From allostatic agents to counterfactual cognisers: Active inference, biological regulation, and the origins of cognition. *Biol. Philos.* **2020**, *35*, 1–45. [CrossRef]
52. Madary, M.; Metzinger, T.K. Real virtuality: A code of ethical conduct. Recommendations for good scientific practice and the consumers of VR-technology. *Front. Robot. AI* **2016**, *3*, 3. [CrossRef]
53. Blackwell, L.; Ellison, N.; Elliott-Deflo, N.; Schwartz, R. Harassment in social virtual reality: Challenges for platform governance. *Proc. Acm Hum. Comput. Interact.* **2019**, *3*, 1–25. [CrossRef]
54. Fallis, D. The Epistemic Threat of Deepfakes. *Philos. Technol.* **2020**, 1–21. [CrossRef]
55. Popper, K. *Conjectures and Refutations: The Growth of Scientific Knowledge*; Routledge: London, UK, 2014.
56. Deutsch, D. *The Beginning of Infinity: Explanations That Transform the World*; Penguin UK: London, UK, 2011.
57. Aliman, N.M.; Kester, L.; Yampolskiy, R. Transdisciplinary AI Observatory—Retrospective Analyses and Future-Oriented Contradistinctions. *arXiv* **2020**, arXiv:2012.02592.
58. Saks, M.J.; Koehler, J.J. The individualization fallacy in forensic science evidence. *Vanderbilt Law Rev.* **2008**, *61*, 199.
59. Miller, D.W. *Out of Error: Further Essays on Critical Rationalism*; Ashgate Publishing, Ltd.: Farnham, UK, 2006.
60. Harding, S. *Can Theories be Refuted? Essays on the Duhem-Quine Thesis*; Springer Science & Business Media: Berlin, Germany, 1975; Volume 81.
61. Illes, M.; Wilson, P. *The Scientific Method in Forensic Science: A Canadian Handbook*; Canadian Scholars' Press: Toronto, ON, Canada, 2020.
62. Hofmann, S.M.; Klotzsche, F.; Mariola, A.; Nikulin, V.V.; Villringer, A.; Gaebler, M. Decoding subjective emotional arousal from EEG during an immersive Virtual Reality experience. *bioRxiv* **2020**. [CrossRef]
63. McDuff, D.; Hurter, C.; Gonzalez-Franco, M. Pulse and vital sign measurement in mixed reality using a HoloLens. In Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology, Gothenburg, Sweden, 8–10 November2017; pp. 1–9.

64. Katie, H.; Zulqarnain, K.; Feldman, M.J.; Catie, N.; Devlin, M.; Dy, J.; Barrett, L.F.; Wormwood, J.B.; Quigley, K.S. *Context-Aware Experience Sampling Reveals the Scale of Variation in Affective Experience*; Scientific Reports; Nature Publisher Group: London, UK, 2020; Volume 10.
65. Ohst, B.; Tuschen-Caffier, B. Does physiological arousal lead to increased catastrophic misinterpretation? An experiment based on the concept of a fear memory. *BMC Psychol.* **2020**, *8*, 1–11. [CrossRef] [PubMed]
66. Renshon, J.; Lee, J.J.; Tingley, D. Physiological arousal and political beliefs. *Political Psychol.* **2015**, *36*, 569–585. [CrossRef]
67. Hao, K. The Biggest Threat of Deepfakes Isn't the Deepfakes Themselves. *MIT Technology Review*, 10 October 2019.
68. Riikonen, A. Decide, Disrupt, Destroy: Information Systems in Great Power Competition with China. *Strateg. Stud. Q.* **2019**, *13*, 122–145.
69. Chittaro, L.; Sioni, R.; Crescentini, C.; Fabbro, F. Mortality salience in virtual reality experiences and its effects on users' attitudes towards risk. *Int. J. Hum. Comput. Stud.* **2017**, *101*, 10–22. [CrossRef]
70. Shehryar, O.; Hunt, D.M. A terror management perspective on the persuasiveness of fear appeals. *J. Consum. Psychol.* **2005**, *15*, 275–287. [CrossRef]
71. Solomon, S.; Greenberg, J.; Pyszczynski, T. A terror management theory of social behavior: The psychological functions of self-esteem and cultural worldviews. In *Advances in Experimental Social Psychology*; Elsevier: Amsterdam, The Netherlands, 1991; Volume 24, pp. 93–159.
72. Woo, G. Downward Counterfactual Search for Extreme Events. *Front. Earth Sci.* **2019**, *7*, 340. [CrossRef]
73. Yoo, J. Ideological Homophily and Echo Chamber Effect in Internet and Social Media. *Stud. Int. J. Res.* **2007**, *4*, 1–7.
74. Jakubowski, G. What's not to like? Social media as information operations force multiplier. *Jt. Force Q.* **2019**, *3*, 8–17.
75. Aliman, N.M.; Kester, L. Transformative AI governance and AI-Empowered ethical enhancement through preemptive simulations. *Delphi* **2019**, *2*, 23–29. [CrossRef]
76. Aliman, N.M.; Kester, L. Artificial creativity augmentation. In Proceedings of the International Conference on Artificial General Intelligence, St. Petersburg, Russia, 16–19 September 2020; pp. 23–33.
77. Caldwell, M.; Andrews, J.; Tanay, T.; Griffin, L. AI-enabled future crime. *Crime Sci.* **2020**, *9*, 1–13. [CrossRef]
78. Sawers, P. The Social Dilemma: How Digital Platforms Pose an Existential Threat to Society. Available online: https://venturebeat.com/2020/09/02/the-social-dilemma-how-digital-platforms-pose-an-existential-threat-to-society/ (accessed on 2 November 2020).