10.2 Cyborgnetic Epistemology and Science

Against the backdrop of the noticeable insufficiency of empiricist epistemologies to get a grip on the epistemic threat landscape of the deepfake era, cyborgnetic epistemology took critical rationalism frameworks as advanced by Popper [404] and reinvigorated by Frederick [198] as point of departure and piecemeal refined those against the epistemically more challenging background of problematic deepfake phenomena. The key epistemic artefact of cyborgnetic epistemology is the phenomenon of new EBs – which are constructed out of explanatory information (EI) blocks (grounded both in language and in physics) that are interconnected in accordance with a rigorously specified epistemic order. While so-called Type I entities (of which all present-day AI systems are a subset) are all those for which it is impossible to understand EI, Type II entities are those for which this is possible. Building on that, a *cyborgnet* is a highly generic substrate-independent term (that is *not* to be confused with the much more narrow concept of a cyborg) and which stands for the template of a dynamic, hierarchical and context-dependent functional unit that can be described by a *directed* graph where EB-based narratives combine *at least* one Type *II* entity with *at least* one Type *I* entity. We describe an intra-cyborgnetic information-theoretical asymmetry between the ability to understand vs. the ability to create information. Due to this so-called *cyborgnetic comprehension bottleneck*, it holds that while it is possible to create all new non-EB-like information x without understanding that information x, it is impossible to create new (i.e. yet unknown) EBs without understanding those. In short, due to the latest developments in Type I AI research, cyborgnetic epistemology was able to directly integrate this factor in its own methodology. In short, cyborgnetic epistemology is itself an act of cyborgnetic creativity augmentation. Strikingly, thanks to the same Type I AI factor, it is also amenable to experimental problematization and is able to enter in and merge with the realm of *science* (see also Chapter 6.1).

To sum up, while in the past the discipline of epistemology was regarded as a widely philosophical pursuit divorced from its object of study, a modern cyborgnetic philosopy of science in the deepfake era becomes epistemically more palpable. In turn, new avenues for experiments are created inserting Type-AI-augmented epistemology in science and Type-I-AI-aided science in epistemology. On the whole, the epistemic aim of cyborgnetic epistemology applicable to all domains of rational reasoning is to create ever better new EBs. Concerning the necessarily updatable criteria for novelty, cyborgnetic epistemology *explicitly* couples it to the forgery abilities of the best state-of-the-art Type I AI. The always relational and thus always comparatively formulated criteria for better EBs are updatable by-design established by agreement requiring no justification (as the latter is logically impossible). Exemplary criteria are e.g. a preference for EBs with more novel problematizable predictions, EBs that are more innovative, more risky, harder-to-vary, bolder or more aesthetically appealing than rival ones. (However, criteria such as “more

trustworthy" are *not* a valid option since highly sensible to manipulation in the deepfake era.) In this way, in line with Popper [404], cyborgnetic epistemology has a preference for *impossibility statements* [344] since those are simultaneously more risky, bolder and harder-to-vary than laxer formulations. The latter is beneficial for science and epistemic security as it allows a faster and more robust piecemeal adaptation to the fastly fluctuating epistemic threat landscape. In short, it avoids an epistemic stagnation in dysfunctional local attractors. Consistent with Frederick [200], it is both rational to pragmatically act in accordance with the currently instated best EBs as it is to act against those. In cyborgnetic epistemology, extending beyond Frederick, a cyborgnet actively integrates Type I AI to both: 1) proactively broaden known old EBs with non-trivial but convergent new non-EB-like EI that can be deduced from currently known old EBs and 2) to generate divergent new non-EB-like EI (which includes noise injection harnessing genuine randomness [84, 245]) that conflicts with known EBs in order to challenge one's own assumptions and unpredictably stimulate one's EB creativity by being able to look around conceptual corners and propagate through mental barriers. In short, cyborgnetic epistemology encourages the conscious *harnessing of stochasticity* [370] by Type-I-AI-augmented cyborgnets to better regulate the epistemically-relevant disorder in the deepfake era. In this way, a cyborgnet uses both genuine randomness and the best EBs to deepen serendipity and broaden creativity such that slow creativity and fast serendipity meet more often.

A further relevant tenet was that next to conjecturing ever better new EBs, the methodology in cyborgnetic epistemology comprises experimental problematization and provisional refutation. An instated EB cannot be (not even temporally) refuted by experimental problematization. Instead, one requires at least one other new EB that is better than that EB in question to provisionally refute it. Given inevitable unintentional (self-)misguidance but also intentional malice to frame epistemic distortion in the deepfake era, it must be epistemically permissible to repeal agreements concerning both the experimental problematization and the refutation of EBs. In this way, a high flexibility is facilitated which still stays rigorous since based on ever better new EBs and not experiments. Importantly, one is *not* attempting to establish whether a candiate new EB is true/truer or wrong/more wrong. This is impossible because truth is related to that undivisible totality, that unanalyzable whole which contains both the cyborgnetically observed (the EB) and the cyborgnetic observer itself. This unanalyzable totality, unknowable as a whole may be linked to what Kant [8] called the noumenon (which is contrasted to the knowable phenomena). One cannot compare one's theories with that Oneness directly. Instead, as part of that totality, one compare's one's theories with one's theory-laden perception of other parts from within that totality. Thus, to recapitulate, in cyborgnetic epistemology, one focuses on whether a new candidate EB is better *in comparison* to the currently best instated EB alternatives and does *not* attempt to ask whether an EB is true/truer, wrong/more wrong. (An EB can also *not* be judged to be "good" in isolation.) As stated by the physicist and philosopher David Bohm [74]: *"If we supposed that theories gave*

true knowledge, corresponding to 'reality as it is', then we would have to conclude that Newtonian theory was true until around 1900, after which it suddenly became false, while relativity and quantum theory suddenly became the truth. Such an absurd conclusion does not arise, however, if we say that all theories are insights, which are neither true nor false but, rather, clear in certain domains, and unclear when extended beyond these domains."

Overall, to sum up, it is thus stated that the goal of epistemology including also in scientific contexts should imply an approach that is EB-anchored, trust-disentangled and adversarial and aims at identifying ever better new EBs. Experimental problematization shapes this epistemic trajectory but does not determine it. Using provisional refutations, EB-anchored science makes pragmatic progress via incremental small steps from old currently best EB to new even better EB, which is why the epistemic aim is of a relational and comparative nature. One can walk forth and back as rationally required. New EBs are universal affordances because one can utilize them to try to better explain the universe as a whole including its genesis. Thereby, the laws of nature that cyborgnets conjecture including the ones that attempt to model the initial conditions of the universe can be formulated or are at least transformable into the format of new EBs at the time they were new. It is thus conceivable that all new EBs *about* the universe as a whole that ever existed, exist now and will exist share a common ground that binds them in a way that they may be non-trivially entangled. Indeed, we share the view of Corazza stating that *"creativity episodes are [...] mutually interconnected through several mechanisms (past and future concatenation, estimation, and exaptation), to form a dynamic universal creativity process (DUCP), the beginning of which can be traced back to the Big Bang of our universe"* [129]. In this sense, note that entities that may initially appear to be disconnected, could have *locally inaccessible* degrees of freedom that would reveal how they are differentially connected in a directed graph hidden "under the hood". For a simple illustration, see Figure 10.1.

In the cyborgnetic DUCP described in Chapter 9.2.1, the space of possible options appears to expand and what was previously considered to be impossible can become accepted to be possible e.g. when a cyborgnet acts against the best EBs instated at a certain point or by cyborgnetic serendipity. Due to that, cyborgnetic epistemology can reach no end state, something that appears clear is highly unstable and may shift conceptually at a later stage. Concerning the metaphor of an *epistemic metamorphosis* from Figure 10.1, note that even the "final" 3D torus perception may not last as it could itself be later perceived by a cyborgnet to itself only correspond to a small part of a much greater figure of higher dimensionality... and so on ad infinitum. Bohm stated that: *"like the processes of nature, those of the mind are basically of an infinite order that is always tending to evolve towards new orders, and thus to develop hierarchies constituting new kinds of structures"* [73]. In Chapter 11, we briefly motivate why for epistemic security reasons, future work could study a new epistemic area that one could call cyborgnetic *epistopology*.

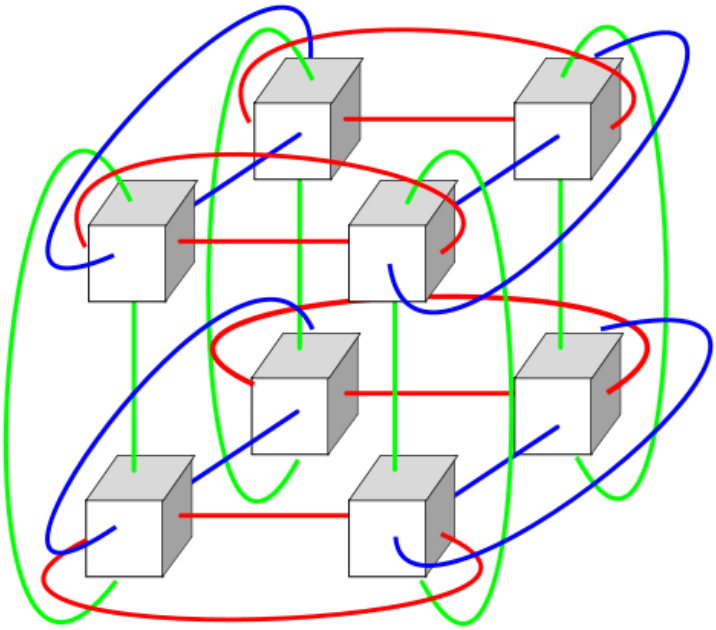

Figure 10.1: Highly simplified illustration for the metaphor of an *epistemic metamorphosis*. Initially, a cyborgnet could be only perceiving separable white *squares* (being 2D facets in this 3D network) and *not* the coloured edges. Then, a 3D structure could emerge mentally e.g. via the shape of a *cuboid*. Suddenly, the previously hidden edges could be understood and a herewith *en*folded 3D *torus* conjectured. Picture taken from [436].

10.2.1 Experimentally Problematizable Impossibility Statements

1. While Type I AI can create new non-EB-like information, including also new non-EB-like EI, it does not understand the latter and it is impossible for Type I AI to reliably create new EBs with arbitrary high accuracy.

2. It is impossible to implement an oracle able to reliably predict the future creation of new EBs itself. In short, an *epistemic perpetuum mobile* is impossible. Creating new EBs comes at the cost of a harder Type-II-only process of understanding which requires cognitive efforts linked to thermodynamical costs.

3. A *moral perpetuum mobile* able to reliably predict the future of all future moral values and norms is impossible because it could imply the creation of new EBs.

4. EB-based rationality *without* core affect[2] (by virtue of being an indispensable continuous ingredient of *consciousness* and mental constructions [49, 54]) is impossible.

[2]Already the criteria for *better* new EBs involve affect. An example is a preference for new EBs that are "more aesthetically appealing" than rival ones. In this connection, Bohm [73] wrote: *"[...] really great scientists have, without exception, all seen in the process of nature a vast harmony of order and indescribable beauty. [...] Indeed, every great scientific theory was in reality founded on such a perception of some very general and fundamental feature of the harmony of nature's order. Such perceptions, when expressed systematically and formally, are called "laws of nature"."*

10.3 AI Design and AI Regulation Recommendations

10.3.1 Mitigating Honey Mind Traps

1. **Avoiding an *over*estimation of present-day AI:** In light of the transdisciplinary knowledge collated in Chapter 9, this could for instance be supported by an education on epistemically-relevant and complexity-related ontological differences: a) non-complex and non-living (such as e.g. a chess software), b) complex but non-living (such as e.g. the Sun), c) living but non-conscious (such as e.g. plants), d) conscious but non-cyborgnetic (such as e.g. birds) and e) cyborgnetic (such as e.g. humans). Presently, all commonly called AI systems are non-conscious. With the exception of e.g. xenobots [68] which are living but non-conscious entities made on the basis of frog cells and which may belong to cluster c), most present-day AI systems belong to cluster a). An epistemically-sensitive AI design would convey to humans that Type I AI from cluster a), b) and c) is *not* conscious. Attempts to fuel attributions of agency and experience would be avoided.
2. **Avoiding an *under*estimation of present-day AI:** An exemplary epistemically-sensitive method would be the conjunction of cyborgnetic creativity augmentation (see Chapter 6.2) and the routine-like integration of that method in the *Co-create* function of a COOCA-loop (see Chapter 8). On the whole, from a design perspective, the goal would be to support the experience of oneness but *not* via the misguided assignment of consciousness to non-conscious Type I AI, but instead by establishing a seamless interaction that is more comparable to the interaction between oneself and language being a Type I tool, between oneself and a new artificial body part or between oneself and an AI-augmented sheet providing new non-EB-like comments on what one writes. Type II agency must be foregrounded by explicitly shifting design narratives from intelligence to EB-based creativity – a process that prohibits global high-risk Type-I-only-loops and where instead, Type I AI becomes part of somebody via a *local* intra-function *encapsulation* within an individual cyborgnetic function of a global cyborgnetic COOCA-loop (see Chapter 8.3 and 8.5.2).

10.3.2 Malicious Deepfake Design Regulation

Any new *non*-EB-like information could be forged (see Section 10.2.1). Old (i.e. already known) EBs could be copied which is traceable and unproblematic. To prohibit specifically new *deepfake* x cannot function in the long-term due to the indistinguishability of new non-EB-like x and new non-EB-like *deepfake* x. One could instead e.g use old laws to regulate any general new manifestation of the old problem x. Because it is impossible to forge new EBs, one does not even *need* to forbid *deepfake* new EBs – they are impossible.