Cyborgnetic Epistemology and Science (more details)

- The epistemic modus operandi is *EB-anchored*, *trust-disentangled* and *adversarial*.
- The epistemic aim is to create ever better new EBs (see example in Figure 6.1).
- A "trust-disentangled" modus operandi signifies that the epistemic modus operandi is grounded in agreed upon criteria for new ever *better* and *not* e.g. "more trustworthy" EBs. This means that it is orthogonal to any trust relation between involved entities. A better EB must be formulated such that metaphorically speaking it appears to defend itself against adversarial candidate EBs.
- An "adversarial" modus operandi signifies a conscious fallibilism at all levels. Firstly, experiments never conclusively falsify a currently accepted (i.e. instated) EB, they make that old EB problematic. We call it *experimental problematization*. Secondly, for a *(provisional) refutation* of an EB, one needs at least a new better EB. All refutations are provisional by design and can be repealed retrospectively (see also Chapter 5). Thirdly, consistent with Frederick [200], it is both rational to act *in accord* with currently instated EBs *and* to experimentally act *against* those. One reason for the latter is that experiments could at any time unexpectedly make even the best tested old EBs problematic. Moreover, to act against old instated EBs could stimulate one's creativity in crafting novel better EBs that refute those.

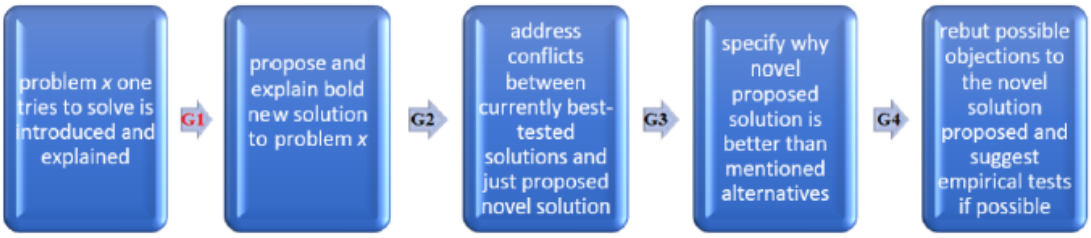

Figure 6.1: Exemplary epistemic total order for the generation of new EBs (the instructions are loosely inspired by an essay of Frederick [199]). Each glue operation x is indicated via a label G_x. EBs are a special form of explanatory information (EI) obtained by interweaving EI blocks via the step-by-step application of rational procedures sampled from a robust explanation-anchored, adversarial and trust-disentangled epistemology. In science, the specification of (direct or indirect) empirical tests in G_4 is the default condition.

- The inherently comparative criteria for "better" EBs are *updatable* and determined by agreement. Current criteria accepted in science encompass e.g. a preference for explanations that are simpler, provide more novel problematizable predictions, are more innovative, more aesthetically appealing than rival ones (see also Chapter 5).

- The inherently comparative criteria for "new" EBs must be udaptable and determined by agreement. As displayed in Figure 6.1, for a candiate EB to be accepted, the novelty criterium must inherenly be fulfilled. Indeed, the glue operation G_1 is formulated as follows: "propose and explain bold *new* solution to problem x". (Note that we must presuppose that problem x is a genuine problem [198] in the first place; not all questions are epistemically-relevant.) In the deepfake era, novelty must be adapted to exclude forgery by even the most advanced Type I AI.

- In a new EB (be it in the science domain or in philosophy), one must specify a *new* solution to a genuine problem x which must fullfil the following two necessary conditions: 1) the solution can be represented as a set of explanations S_E *and* 2) that set S_E contains *at least one* explanation $e_{Mysterious}$ that could *not* have been *reliably* generated with arbitrary high accuracy via an automatable (i.e. Type-*I-only*) pipeline given existing knowledge. The latter implies the following: given publicly available knowledge S_{OldEBs} and the genuine problem x as inputs, it would be impossible for a Type I AI to reliably generate $e_{Mysterious}$ as output.

- In the science domain, a new EB must additionally fullfil the following third and fourth necessary conditions. The third necessary condition is that the set S_E entails *at least one* new prediction $p_{Mysterious}$ for which it holds that: a) it is in theory amenable to experimental problematization but has not yet been made problematic by experiment in practice *and* b) it could *not* have been *reliably* generated with arbitrary high accuracy via an automatable (i.e. Type-*I-only*) pipeline given the set S_{OldEBs} of currently instated old EBs. This implies the following: given publicly available knowledge S_{OldEBs}, it would be impossible for a Type I AI to reliably generate $p_{Mysterious}$. Finally, consistent with Frederick [198], the fourth necessary condition is that this prediction $p_{Mysterious}$ could *not* have been deduced *without* combining *all* elements from the set of explanations S_E.

- In the philosophy domain, one can accommodate for such lines of thought by specifying that a new EB must also fulfill the necessary condition that it is currently *not* possible to identify a subset S_{SubE} such that $S_{SubE} \subsetneq S_E$ with S_{SubE} being already *sufficient* to solve the problem x. The latter avoids superfluous statements.

- While it holds that 1) Type I AIs can in theory forge the creation of any new non-EB-like information including texts widely perceived by humans as "novel explanations", it holds that 2) due to a gap of understanding, it is impossible for all Type I entities to reliably create new yet unknown EBs respecting an epistemic total order stemming from a rigorous epistemology as e.g. exemplified in Figure 6.1.

- An experimental problematization of cyborgnetic epistemology would e.g. be a shortcut via a Type *I* AI able to reliably create new EBs with arbitrary high accuracy.
- A (provisional) refutation of cyborgnetic epistemology would be a better new theory that explains why such a Type-I-shortcut is possible.