
EMERGENCE AND EVIDENCE: A CLOSE LOOK AT BUNGE'S PHILOSOPHY OF MEDICINE

Rainer J. Klement^{1*} and Prasanta S. Bandyopadhyay²

¹ Department of Radiation Oncology, Leopoldina Hospital Schweinfurt, Robert-Koch-Straße 10, 97422 Schweinfurt, Germany; rainer_klement@gmx.de

² Department of History & Philosophy, Montana State University, Bozeman, MT, USA; psb@montana.edu

* Correspondence: rainer_klement@gmx.de; Tel.: +49 9721 7202761

ABSTRACT

In his book “Medical Philosophy: Conceptual issues in Medicine“, Mario Bunge provides a unique account of medical philosophy that is deeply rooted in a realist ontology he calls “systemism”. According to systemism, the world consists of systems and their parts, and systems possess emergent properties that their parts lack. Events within systems may form causes and effects that are constantly conjoined via particular mechanisms. Bunge supports the views of the evidence-based medicine movement that randomized controlled trials (RCTs) provide the best evidence to establish the truth of causal hypothesis; in fact, he argues that *only* RCTs have this ability. Here we argue that Bunge neglects the important feature of patients being open systems which are in steady interaction with their environment. We show that accepting this feature leads to counter-intuitive consequences for his account of medical hypothesis testing. In particular, we point out that (i) the confirmation of hypotheses is inherently stochastic and affords a probabilistic account of both confirmation and evidence which we provide here; (ii) RCTs are neither necessary nor sufficient to establish the truth of a causal claim; (iii) testing of causal hypotheses requires taking into account background knowledge and the context within which an intervention is applied. We conclude that there is no “best” research methodology in medicine, but that different methodologies should co-exist in a complementary fashion.

Keywords: Bayesianism; confirmation; evidence; evidence-based medicine; Mario Bunge; systemism; systems thinking; philosophy of medicine; philosophy of science

1. Introduction

In his book „Medical Philosophy: Conceptual issues in Medicine“ the scientist and philosopher Mario Bunge contrasts what he calls “scientific medicine” with complementary and alternative medicine (CAM)¹ which according to him “is a broad panoply of therapies lacking in both scientific basis (knowledge of mechanism) and evidence (randomized controlled trials)” [1]. His distinction between these two types of medicine is not only epistemological (according to him only scientific medicine has knowledge of mechanisms and is evidence-based), but also an ontological one (that

¹ Bunge makes no further distinction between traditional, complementary and alternative medicine which is problematic, since all three refer to different concepts of medical systems. However, a detailed criticism of this conflation is not our concern here.

refers to the some features of the world): While he perceives CAM embedded in a holistic worldview (“the hole precedes and dominates its parts”), scientific medicine “treats patients like ... systems that can be dismantled, at least conceptually, with the help of modern biology: it is systemic.”

Bunge’s ontological characterization of modern medicine as systemic follows his earlier work on the philosophy of social science. It is not found within other books on the philosophy of medicine [2–5]. It therefore deserves a critical appraisal, in particular concerning its implications for epistemology and medical research praxis (methodology). This is what we aim for in this paper. We are thereby not concerned with his philosophy of medicine as such, but with the broader implications that emerge from his realist-systemic ontology concerning emergence of new properties not reducible to some more fundamental properties and evidence. We agree that such an ontology seems indeed quite reasonable given the successful history of systems thinking in biology and its increasing applications in medicine [6–8]. However, there is a lack of meta-research on the implications of systems thinking for the hallmarks of modern medicine such as randomized controlled trials (RCTs) and evidence-based medicine (EBM) in general.

In the next section, we first provide a motivation for systems thinking and anti-reductionism in medicine and then describe Bunge’s account of systemism in medicine that adopts the Humean conception of causality in which cause and effect are constantly conjoined. In section 3, we will review Bunge’s epistemological principles for good scientific practice in medicine, and relate it to the current paradigm of EBM. We are going to highlight that both Bunge and EBM emphasize a hierarchical structure of medical evidence and research methodologies, with RCTs conceived as providing the highest quality evidence for establishing the truth of medical hypotheses. In section 4 we are going to develop our critique of Bunge’s conception of causality based on the argument that humans are not closed, but open systems which are in steady interaction with their environment. The acceptance of this fact is at odds with both a constant conjunction of events and the proposition of a hierarchy of methodologies with RCTs on top. Instead, we argue for a “circular” or integrative view of evidence in which multiple methodologies co-exist with each other and the best method is dependent on the particular context and research question.

2. Systemism, emergence and medicine

2.1. Background

The view that every medical phenomenon can be explained by referring to entities at a lower ontological level, also called entity reductionism, appears to be the default view in medicine [3, chapter 4]. Examples abound: Diseases are reduced to their syndromes which are treated by pharmaceutical drugs; individual biomarkers which lie outside their normal range are “corrected” without considering system-wide effects [6]; tumors are mostly conceived as independent entities and ultimately reduced to the level of oncogenic driver mutations or tumor suppressor loss-of-function mutations for which targeted therapies could be designed – which is then called “precision medicine”; humans with a disease are reduced to patients instead of also being agents (capable of taking actions by themselves) [9]. Finally, a reductionist view may also be held responsible for the “junk science” produced within the field of nutrition research where dietary complexity and that of the person or animal eating the diet are all too often reduced to associations of single macro- and micronutrients with certain outcomes, leading to false and sometimes counter-intuitive conclusions [10,11].

Such reductionism in medicine is challenged from a biological perspective considering that macroscopic multicellular life possesses a kind of complexity that cannot be solely explained through the properties of its constituent parts. Bains and Schulze-Makuch for example, by comparing a fruit fly to an equivalent mass of *Escherichia coli* (*E. coli*), write:

The mass of *E. coli* can be described by describing one *E. coli* and then saying “grow 10^9 of them”, [while] describing a fruit fly requires describing all of its cell types and their interactions in chemistry, space and time [12].

This example highlights Bunge’s conception of *emergence*. The fruit fly possesses certain properties that none of its constituents possesses, e.g. sexuality. These are *emergent* properties [13]. In contrast, the conglomerate of *E. coli* bacteria lacks any special property that would not also be possessed by an individual *E. coli* bacterium, and hence has no emergent properties. More generally, Bunge defines emergent properties as systemic properties that none of the parts of the system share, but that are explainable by the parts of the system *and* its entire organization [13]. In other words, according to him, certain systemic properties are ontologically emergent properties, and are not anyway reducible to some more basic properties. Insofar, he is an ontological anti-reductionist. He thinks, however, that every system with emergent properties can be explained by our knowledge about its lower-level properties and their connections. Thus, he is an epistemological reductionist. Bunge calls his stance *rational emergentism*. In his view, every aspect of the fruit fly within a given context should be principally explainable if all of the fly’s constituents plus all their interactions within this context would be known. Since diseases can be conceptualized as emerging properties of individual patients, scientific medicine should aim at understanding (i) the underlying properties that give rise to a disease; (ii) their interaction with the environment; (iii) their manipulation through medical interventions. Our argument, developed in more detail below, will be that the second point is strongly neglected in Bunge’s account of medical hypothesis testing.

2.2. Bunge’s ontological systemism

Bunge’s systemism is influenced by Paul Henri Thiry d’Holbach’s *Système de la Nature* (published in 1770) and Ludwig von Bertalanffy’s *General System Theory* [14]. For Bunge, the world is an objective reality made up of systems or parts thereof and therefore stratified into different levels or “layered”. This is the ontology he refers to as “systemism”. In more detail, his systemic worldview rests on the following two postulates, stated in an earlier paper [15]:

(S1) *Everything, whether concrete or abstract, is a system or an actual or potential component of a system;*

(S2) *Systems have systemic (emergent) features that their components lack.*

Bunge argues that “systemism is the only ontology that fits the modern sciences” and that, “because it entails emergentism, systemism overcomes reductionism in its various forms” (p.15).² He emphasizes that systemism entails ontological emergence since systems are postulated to possess properties that only emerge from the interaction of their parts. As examples for such emergent properties within medicine, Bunge names apoptosis on the level of cells (p. 138) or being alive, thinking and socializing (p.16) on the level of the whole system.

Bunge classifies systems broadly into being concrete/material (stars, cells, humans, social systems etc.) or abstract (properties, events, ideas, feelings etc.). Concerning the former he writes [1, pp.13-14]:

The conceptual or empirical analysis of a concrete system, from atom to body to society, consists in identifying its composition, environment, structure, and mechanism. These components may be schematically defined as follows:

Composition = Set of constituents on a given level (molecular, cellular, etc.)

Environment = Immediate surrounding (family, workplace etc.)

² Bunge speaks of emergentism instead of emergence as in his former publications.

Structure = Set of bonds among the components (ligaments, hormonal signals, etc.)

Mechanism = Process(es) that maintain(s) the system as such (cell division, metabolism, circulation of the blood, etc.)

Bunge defines the environment of a system as the set of all entities not being part of the system, but interacting with it. He generally conceives causal mechanisms as processes (sequences of events) in concrete systems involving energy transfer that are activated by particular events (the causes) [16,17]. Mechanisms can act between different levels of a system in both top-down and bottom-up direction (Figure 1). Bunge adopts the Humean view of causation – event X is a cause of event Y if and only if (i) X is both a necessary and sufficient condition for Y to happen, and (ii) X happens before or simultaneously with Y. In this sense, mechanisms act as mediators between causes and their effects.

While Bunge mentions the importance of the (social) environment and its interaction with patients, he makes no distinction between open and closed systems in his medical philosophy. This is despite the concept of an open system as “the characteristic state of the living organism” having been introduced by von Bertalanffy [14] who is cited as an influencer of systemism in Bunge’s book. According to von Bertalanffy, an open system is characterized by a “steady inflow and outflow of materials” [14]. Contrary to closed systems, open systems therefore transcend both conventional thermodynamics and one-way causality of the form “this one event always causes this or that effect” [18]. We more generally conceive an open system as a system which is in steady interactions with its environment, whereby these interactions can give rise to events and the triggering or avoidance of mechanisms; they hence influence causal effects. As we will argue in section 4, Bunge’s failure to clearly characterize patients as open systems is the major flaw in Bunge’s medical philosophy, giving rise to several epistemological and methodological problems in his account.

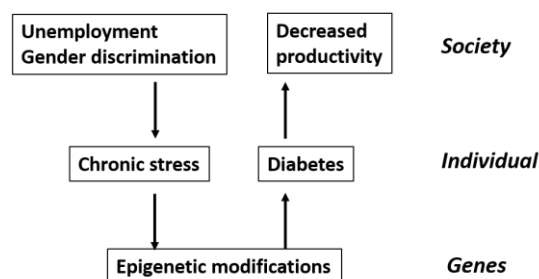


Figure 1. Two examples for mechanisms acting across levels in top-down or bottom-up direction. In the former case, macrosocial factors such as unemployment or gender discrimination induce chronic stress in affected individuals which in turn negatively affects gene transcription through epigenetic modifications. Such modifications of gene transcription can promote the development of systemic diseases such as (type II) diabetes which in turn has negative effects on the social level by decreasing productivity or inflicting additional costs to the health care system. These examples are taken from page 35 of Bunge’s book [1].

3. Bunge’s epistemology

Bunge’s epistemological principles for scientific medicine can best be summarized by three further postulates that he stated in a previous paper along with the two ontological postulates cited in the previous section [15]:

(S3) *All problems should be approached in a systemic rather than in a sectorial fashion;*

(S4) *All ideas should be put together into systems (theories); and*

(S5) *The testing of anything, whether idea or artifact, assumes the validity of other items, which are taken as benchmarks, at least for the time being.*

According to Bunge, and in line with several other philosophers of science, biomedical research is particularly concerned with the testing of causal hypotheses³ Causal hypotheses require causal mechanisms which may be conjectured based on statistical data (e.g. observational studies) [1,16]. Bunge has a Humean conception of causality in which cause and effect are constantly conjoined. Such conception was once the standard view in medicine. It was first questioned in the 1950s when the problem whether smoking causes lung cancer was being investigated more thoroughly [5]. Smoking is neither necessary (asbestos, e.g., can also cause lung cancer) nor sufficient (not every smoker gets lung cancer) for lung carcinogenesis. This example clearly showed that there could be causes that are not always followed by their effects; it required a new concept of causality called indeterministic or probabilistic causality and spurred new philosophical investigations of the link between causality and probability that last until today [5]. Bunge, however, rejects any fundamental connection between causation and probability. On Bunge's account, probability measures the objective possibility or likelihood of an individual fact, but only when there is chance involved, either within things themselves (atoms) or the sampling procedure that determines the observation of the facts (randomization). He therefore rejects Bayesianism in its broadest sense which he accuses of being subjective, absurd and even dangerous when applied to diseases and therapy [1, p. 101].⁴ We are going to discuss the AIDS example which he uses to justify this claim in the next section.

The adoption of causal determinism and abandonment of probabilities⁵ characterize his account of medical hypothesis testing and his concepts of evidence and confirmation. For Bunge, it is necessary to subject medical hypotheses to tests of two kinds – a *conceptual test* according to his postulate (S5), evaluating whether the hypothesis is consistent with the bulk of prior knowledge; and an *experimental test* that can only be conducted through a randomized controlled trial (RCT). Bunge conceives randomization into a treatment and control group necessary to correct for biases that are due to the heterogeneity of human individuals. Only randomization could guarantee an equal distribution of all unknown variables with a possible influence on the outcome between treatment and control group, “so that the intervention stand out as the *only cause* of the difference in the outcome of the study” [1, p. 142].

This argument echoes the one made by the evidence-based medicine (EBM) movement that has always considered RCTs as the “gold standard” providing the best available evidence, because only RCTs would theoretically balance any unknown factors between treatment and control group confounding the association between an intervention (the cause) and its effect. Hence, RCTs (and their meta-analyses) were conventionally placed on top of the so-called “evidence hierarchy”, in which non-randomized trials and observational studies are graded with much lower evidence quality [19]. We agree with EBM proponents who claim that “the higher the quality of evidence, the closer to the truth are estimates of diagnostic test properties, prognosis, and the effects of health interventions” [19]; however, if high quality evidence is set equivalent to particular quantitative research methods generating the evidence as suggested by various EBM “evidence hierarchies”, this amounts to embracing a positivist-empiricist philosophy.⁶ Such reasoning has evoked much criticism from both medical practitioners [20–23], methodologists and philosophers of science [24–30].

³ He broadly distinguishes three kinds of causal hypotheses: (i) Null hypotheses of no association between two putatively causally connected variables. (ii) General hypotheses of the form “X is a cause of that disease” or “the mechanism of X is Y”. (iii) Particular hypotheses such as “that individual is likely to suffer from that disease”.

⁴ This sounds like an argument by definition. Bunge postulates a meaning for “probability” and then concludes that the Bayesian conception is absurd. However, we are not going to pursue this point further here.

⁵ Bunge seems to have missed that Bayesians of the personalist stripe are ontological determinists and epistemic probabilists.

⁶ We use the term positivist-empiricist referring to a conjunction of concepts of logical positivism and Humean empiricism. The former rejects any reference to unobservable (metaphysical) entities, while the latter describes

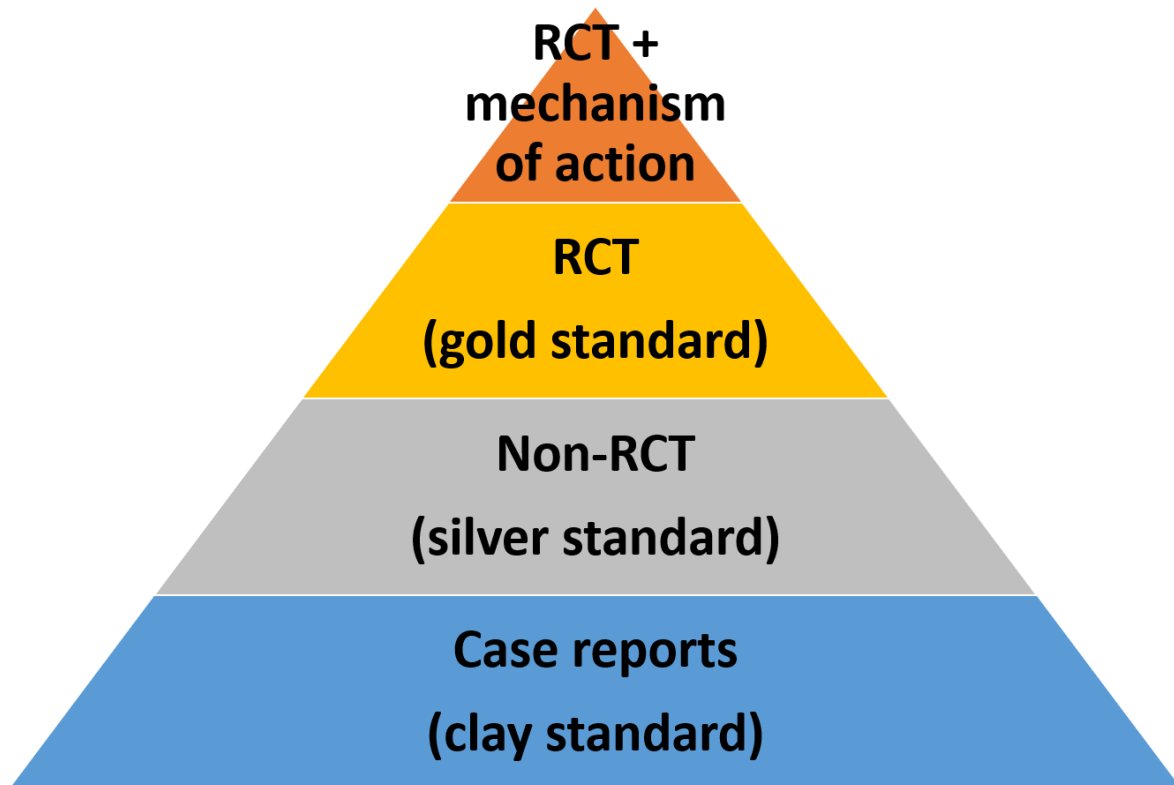


Figure 2. Bunge’s “biomedical research rigor” hierarchy. Such a pyramid is the stereotype of an “evidence hierarchy” widely promoted especially in the early phases of the EBM movement [19].

Bunge rejects most of the criticisms against the evidence hierarchy and “uphold[s] the ruling opinion that the RCT is the gold standard of biomedical research [since] it is the most objective and impartial, hence also the most reliable and responsible, method for assessing the effectiveness of medical interventions on members of heterogenous populations” [1,page 149]. According to Bunge, RCTs are necessary, although not sufficient, to establish the truth of a medical hypothesis. They are insufficient because they overlook *mechanisms of action*. Therefore, Bunge proposes a new pinnacle of the methodological hierarchy (the “platinum standard”) consisting of (i) double-blind and placebo-controlled RTCs with (ii) the explicit statement of a mechanism of action which can account for the observed associations (Figure 2). The order below his supposed platinum standard is analogous to the traditional EBM evidence pyramid in that it ranks evidence from RCTs higher than that from non-RCTs (e.g. case-control or cohort studies) which in turn deliver evidence that is ranked higher than that from case series and case reports. Through his emphasis on evidence for mechanisms contributing to the platinum standard, Bunge departs from EBM in which studies providing evidence for mechanisms are ascribed the lowest quality at the bottom of the various evidence hierarchies [31]. In this way he attempts a rationalism-empiricism synthesis that “is fruitful only jointly with realism” [1, p. 81]. However, Bunge’s views on the epistemological merits of research methodologies below RCTs appear even more extreme than those of EBM; he simply denies that they have any power for testing causal hypotheses:

[O]nly RCTs allow researchers to find out whether a medical treatment is effective. (If preferred, this method is used to find out whether propositions of the form “This therapy is effective” are true or false.) [1, p. 143]

the scientific endeavor to test causal hypotheses by finding quantitative associations between observed events. Indeed, the literature on the philosophical and methodological foundations of EBM emphasizes statistical methodologies and avoids any reference to a particular ontology [2,19].

Bunge uses the word “true” referring to the concept of factual truth that he sharply distinguishes from probability. For him, truth is a property of propositions such as hypotheses, while probability is a property of *individual* random facts (“things and events out there”, p. 99) and therefore ontological. Truth values measure the degree to which a hypothesis corresponds to the facts and are measured by conducting experiments. In contrast, probability values are theoretically assumed, guessed or calculated by frequencies, but never directly measured. Hence, it is not the goal of RCTs to estimate probabilities of individual facts, but rather to estimate truth values of hypotheses. Concerning the assignment of such truth values to scientific hypotheses Bunge writes:

The aim of an experiment, contrary to that of an observation or a measurement, is to garner empirical data relevant to a hypothesis, to test it and find out its degree of factual (or empirical) truth (true, true within such an error, or false). When there is a theory (hypothetico-deductive system of propositions) referring to the same facts, the hypothesis that undergoes an empirical test can also be assigned a theoretical truth value. In both cases, the truth in question is factual, not formal or mathematical [1, page 131].

Therefore, according to Bunge, it is empirical data stemming from an experimental study and being relevant to a hypothesis that are necessary for both testing the hypothesis and assigning it a truth value. Such data then become evidence:

A datum becomes evidence when confronted with a hypothesis: in this case it either confirms or weakens the hypothesis to some extent [1, page 28].

Thus, in Bunge’s view data that are relevant to a hypothesis are evidence for the hypothesis and simultaneously confirm/disconfirm and assigns a truth value to it. However, Bunge gives no unequivocal answer to the question of how exactly a causal hypothesis is supposed to be confirmed or refuted. On one occasion he claims that “[o]nly the unveiling of a mechanism can confirm or refute a causal guess” (p. 100), while on another occasion he states that “confirming or refuting a causal hypothesis, is what is gained when one subjects it to a controlled experiment” (p. 155). We will show in section 4 how the important role of both mechanisms and statistical data for confirming a causal hypothesis follows naturally from a Bayesian analysis of causal hypotheses. We also show how a conceptual distinction between evidence and confirmation is helpful to understand scientific inference including inference in the medical sciences. In contrast to Bunge’s description of medical hypothesis testing, our concepts of evidence and confirmation are able to account for prior knowledge, uncertainties in inference and the objective comparison between different hypotheses based on observed data.

4. A critique of Bunge’s medical philosophy

Having reviewed the ontological and epistemological principles of Bunge’s medical philosophy, which heavily rests on his general account of systemism, we are now going to formulate our major critique against Bunge’s account:

Humans are open systems ⇒

(C1) The confirmation of hypotheses is inherently stochastic and must be distinguished from the notion of objective truth; only comparisons between hypotheses in light of background knowledge and their ability to explain the observed data are objective

(C2) RCTs are neither necessary nor sufficient to establish the truth of a causal claim

(C3) Testing of causal hypotheses requires taking into account background knowledge and the context within which an intervention is applied

In the following three subsections we will justify our claims (C1) to (C3) in more detail.

4.1. Evidence and confirmation as separate concepts

According to von Bertalanffy, who is cited in Bunge's book as a pioneer of systemism, every living system is an open system [14,18]. In system terms, open systems are characterized by steady interactions between the different components of the systems, the systems as a whole with their components and the systems and their components with other systems and their components in the environment. These interactions would give rise to events and hence causality, because interactions represent causal relationships between events; however, they are so complex and dynamic that causality cannot be considered as a constant conjunction of events. That would only be the case in closed systems, something that can be only be established artificially through particular experimental setups in some natural sciences such as physics. In natural open systems, however, causality arises from a tendency on behalf of the system to produce certain patterns or regularities under particular contexts [32]. As a consequence, there are no universal regularities of the form "whenever event X, then event Y", only what appear as such *on average* and what Lawson has named "demi-regularities". Conceptualizing humans as open systems, it becomes clear that the same medical intervention applied in different study settings not always leads to the same outcome since the outcome depends on the environment/ context in which it occurs. It is crucial to point out that statistical methodologies not only presuppose, but only work in closed systems if the goal is to establish causality [33]. One must therefore accept that variations in regularity are predicted to occur in experimental studies in biology, medicine and sociology. This will require the utilization of probabilities:

Variations in regularity are generally specified probabilistically or stochastically, as random processes occurring in the ontic domain. Probability is a measure of the likelihood of an event occurring. The re-conceptualization of stochastic event regularities using the concepts of probability, might be styled 'whenever event x, then on average event y' [33].

For realists, the truth of causal hypotheses cannot be established in an objective way through statistical data alone due to the unavoidable limitations of experiments conducted on biological open systems, or in other words the impossibility to achieve complete closure of a biological open system in order to nail down the true causal effect of an intervention. It follows that the confirmation or disconfirmation of a hypothesis by statistical data is not about assigning (objective) *truth values* as Bunge claims, but about raising or lowering an agent's (subjective) *belief in the truth* of the hypothesis. Once framed, a realist will seek to scrutinize a causal hypothesis in further tests which hopefully provide stronger and stronger confirmation of it [32]. At the same time, the realist will consider different competing hypotheses/models about the data-generating causal processes that she attributes to different entities that are or may be real; the data may then decide between these hypotheses in an objective way.

We have developed two distinct Bayesian accounts to capture these two concepts about the testing of statistical hypotheses [34,35]. The first is an account of belief/confirmation, the second of evidence. Many Bayesians interpret confirmation relations in various ways. For us, an account of confirmation explicates a relation, $C(D,H,B)$ among data D , hypothesis H , and the agent's background knowledge B . For Bayesians, degrees of belief need to be fine-grained. A satisfactory Bayesian account of confirmation, according to us, should be able to capture this notion of degree of belief. In formal terms:

D confirms H to some degree if and only if $P(H|D) > P(H)$

The posterior/prior probability of H could vary between 0 and 1. Confirmation becomes strong or weak depending on how great the difference is between the posterior probability, $P(H|D)$, and the prior probability of the hypothesis, $P(H)$. $P(H|D)$ represents an agent's degree of belief in the hypothesis

after the data are accumulated.⁷ $P(H)$ stands for an agent's degree of belief in the hypothesis before the data for the hypothesis have been acquired. The likelihood function, $P(D|H)$, provides an answer to the question "how likely are the data given the hypothesis"? $P(D)$ is the marginal probability of the data averaged over the hypothesis being true or false. The relationships between these terms, $P(H|D)$, $P(H)$, and $P(D|H)$, and $P(D)$ are succinctly captured in Bayes' theorem:

$$P(H|D) = P(D|H) \times P(H) / P(D) > 0.$$

While this account of confirmation is concerned with belief in the truth of a single hypothesis, our account of evidence compares the merits of two hypotheses, H_1 and H_2 (which could be $\neg H_1$) relative to the data D , auxiliaries A , and background information B . We conceive the evidence of one hypothesis versus the other as an objective function of the data generating process⁸, which takes place via observed or unobserved mechanisms within the system under study. This is also how Bandyopadhyay, Brittan and Taper interpret the likelihood that determines the evidence [35, p.30]:

It is natural to assume that the "propensity" of a model to generate a particular sort or set of data represents a causal tendency on the part of natural objects being modeled to have particular properties or behavioral patterns and this tendency or "causal power" is both represented and explained by a corresponding hypothesis.

As such, evidence provides the link between "the Real" about which we construct hypotheses and "the Empirical" which we observe as patterns or regularities.⁹ Our concept of evidence is therefore consistent with a realist-systemic ontology. Note that this concept also fulfills Bunge's postulate (S5) by explicitly taking background knowledge into account. Such background knowledge and auxiliaries allow deriving evidence through a variety of methodologies, as long as the data are relevant to an aspect of the hypotheses being compared. For example, observing a high correlation between treatment X and effect Y in a RCT may *in theory* provide the strongest evidence for the claim that X causes Y when the alternative is that X is no direct cause of Y , but X and Y are correlated because both are caused by some third (confounding) factor. In contrast, a single case report of a patient taking a drug and developing a serious side effect together with background knowledge about the biological actions of the drug may provide strong evidence for the hypothesis that the drug is harmful in particular contexts.¹⁰ Finally, preclinical *in vitro* and *in vivo* studies may provide strong evidence in favor of a particular mechanism underlying an observed correlation between treatment and outcome.

Because evidence is not a belief relation, but a likelihood ratio, it need not satisfy the probability calculus. The data D constitute evidence for $H_1 \& A_1 \& B$ against $H_2 \& A_2 \& B$ if and only if

$$[P(D|H_1, A_1 \& B) / P(D|H_2, A_2 \& B)] > 1.$$

Bayesians use the Bayes factor (BF) to make this comparison, while others use the likelihood ratio (LR) or other functions designed to measure evidence. For simple statistical hypotheses with no

⁷ More precisely, we can speak of the degree of belief in the truth of the hypothesis; this unifies Bunge's arguments about hypotheses being confirmed and hypotheses being assigned truth values.

⁸ The data generating process is covered by the auxiliary hypotheses which can be about mechanisms. These auxiliaries serve as links between theoretical entities of the system under study and observable features in nature, in this way generating observable predictions [51].

⁹ The distinction between the Real and the Empirical is borrowed from Critical Realism which itself exhibits many features of systems thinking [50].

¹⁰ An example is a recent case report of a patient with chronic kidney disease experiencing severe lactic acidosis from taking the widely prescribed anti-diabetic drug metformin [52]. In the case report, the authors state their background knowledge as follows: "In the setting of dehydration with resultant acute kidney injury, metformin can accumulate, leading to type B lactic acidosis, especially in the presence of other nephrotoxic agents (ACEi and loop diuretics)". They conclude to "use this patient as an example of the population that actually needs dosing adjustments", a theoretical generalization supported by the evidence obtained from observations on this single patient.

free parameters, the Bayes factor and the likelihood ratio are identical, and capture the bare essentials of an account of evidence without any appeal to prior probability. However, the LR becomes an inadequate measure of evidence whenever there are free parameters to estimate; the greater the number of parameters, the more biased the LR becomes. This is what information criteria such as AIC or BIC try to account for [36]. For hypotheses under which there are unknown parameters θ , the densities¹¹ $P(D|H,A\&B)$ are obtained by integrating over the parameter space, so that

$$P(D|H,A\&B) = \int P(D|\theta,H,A\&B)\pi(\theta|H,A\&B)d\theta \text{ [37].}$$

An immediate corollary of the evidential condition (E) is that there is equal evidential support for both hypotheses only when $BF=1$ (or $LR=1$). The numerical value of the BF or LR which distinguishes weak from strong evidence for H_1 versus H_2 is determined contextually and may vary depending on the nature of the problem. It also follows that evidence is accompanied by confirmation and vice versa in the special case that two hypotheses are mutually exclusive and jointly exhaustive. In this case, if the data provide evidential support for H against $\neg H$, i.e. $P(D|H) > P(D|\neg H)$, then it follows from Bayes' theorem that $P(H|D) > P(H)$. However, even in this case a hypothesis for which the evidence is very strong may not be very well confirmed while a claim that is very well confirmed may have no more than weak evidence going for it [35, p.38]. Finally, we note that in many scientific studies, no precise quantitative determination of likelihoods, priors and posteriors of hypotheses might be possible. Even then our concepts remain useful for making qualitative or comparative statements about hypotheses. For example, a qualitative evidential statement may be "the data provide more/equal/less evidence for H_1 compared to H_2 "; a comparative statement relating to confirmation may be " H_1 is better confirmed/equally confirmed/less confirmed by the data than H_2 ".¹²

We will now demonstrate the usefulness of our confirmation/evidence distinction using an example provided by Bunge himself, the purpose of which was supposed to reject Bayesianism as unreasonable [1]:

It is well known that HIV infection is a necessary cause of AIDS: no HIV, no AIDS. In other words, having AIDS implies having HIV, though not the converse. Suppose now that a given individual b has been proved to be HIV-positive. A Bayesian will ask what is the probability that b has or will eventually develop AIDS. To answer this question, the Bayesian assumes that the Bayes' theorem applies, and writes down this formula: $P(AIDS|HIV)=P(HIV|AIDS) \cdot P(AIDS)/P(HIV)$, where an expression of the form $P(A)$ means the absolute (or prior) probability of A in the given population, whereas $P(A|B)$ is read (or interpreted) as "the conditional probability of A given (or assuming) B ."

If the lab analysis shows that b carries the HIV, the Bayesian will set $P(HIV)=1$. And, since all AIDS patients are HIV carriers, he will also set $P(HIV|AIDS) = 1$. Substituting these values into Bayes' formula yields $P(AIDS|HIV) = P(AIDS)$. But this result is false, since there are persons with HIV but no AIDS. What is the source of this error? It comes from assuming tacitly that carrying HIV and suffering from AIDS are random facts, hence subject to probability theory. The HIV-AIDS connection is causal, not casual; HIV infection is only a necessary cause of AIDS. In conclusion, contrary to what Bayesians (and rational-choice theorists) assume, it is wrong to assign probabilities to all facts. Only random facts, as well as facts picked at random, have probabilities.

¹¹ These probabilities of the data are also known as marginal or integrated likelihoods; some authors also denote them as "evidence" [e.g., 53,54], which must not be confused with our account of evidence that always implies a comparison between *two* competing simple statistical hypotheses.

¹² See Klement [55] for such a qualitative medical application of the evidence/confirmation distinction.

The example is supposed to show a paradox arising from Bayesian reasoning. The paradox is that a positive HIV test result provides no confirmation for the hypothesis AIDS, i.e., that b has or will develop AIDS, since the posterior probability of AIDS after obtaining a positive test result is the same as its prior probability. However, the paradox only arises because Bunge is wrong in two assumptions: First, that a positive test result is “true”, so that $P(\text{HIV})=1$, and second that the test has perfect sensitivity, so that $P(\text{HIV}|\text{AIDS})=1$. Both assumptions are at odds with realistic assumptions about tests on open systems which are never perfect. Regarding Bunge’s first assumption, he mistakenly identifies “knowing or observing the data” with “the probability of the data” [35, p.137]. For a Bayesian realist the positive test result is the realization of some data generating mechanisms (in this case, mechanisms of the disease AIDS) modelled by a binary random variable taking on the value of either 0 or 1, so that the correct way of writing $P(\text{HIV})$ is

$$P(\text{HIV}=1)=\sum P(\text{HIV}=1|H_i)P(H_i)=P(\text{HIV}=1|\text{AIDS})P(\text{AIDS})+P(\text{HIV}=1|\neg\text{AIDS})P(\neg\text{AIDS}).$$

This expression includes both the true positive rate (sensitivity) and false positive rate (1-specificity) none of which are exactly 100% or 0%, respectively, in medical tests. In this specific example, the assumptions $P(\text{HIV}|\text{AIDS})\approx 1$ and $P(\text{HIV}|\neg\text{AIDS})\approx 0$ can indeed be justified based on generally very high sensitivity and specificity of HIV tests (although there is clear variation in these test performances across different settings [38], emphasizing the importance of environment/context). However, the prior probability of b having or not having AIDS before the test result is known is also important. In general, observing HIV is the case does not therefore imply $P(\text{HIV})=1$.

On our account, a positive test indeed provides strong evidence for the AIDS hypothesis because $P(\text{HIV}|\text{AIDS})\gg P(\text{HIV}|\neg\text{AIDS})$. In accordance with our intuition, this does not depend on our prior beliefs about the person having or not having AIDS in the first place. What we should believe about b having AIDS after the positive test result has been obtained is however a different question, and again in accordance with our intuition, the answer to this question should now depend on the context, e.g. what we know about the individual and its social relationships. Bunge is not able to capture these intuitions. On his account, solving the inverse problem of going from the results of a medical test or some sign S of a disease D to the precise diagnosis of D can only be achieved if there is a single mechanism M that when conjoined with S is necessary and sufficient for D to occur. In the AIDS example above his reasoning goes as follows: *AIDS occurs* \Leftrightarrow *HIV infection & slow immune reaction*, where slow immune reaction describes the mechanism by which HIV leads to immune system failure. More generally his reasoning is (p. 88):

$$\textit{For all } x: (Dx \Leftrightarrow Mx) \ \& \ \textit{For all } x: (Mx \Leftrightarrow Sx) \ \therefore \ \textit{For all } x: (Dx \Leftrightarrow Sx)$$

Bunge’s solution presupposes that the mechanisms causally linking the signs and the disease always operate the same way, regardless of the context. In other words, he presupposes a closed system, which is not even approximately the case given that medical tests have sensitivities and/or specificities varying across contexts and often less than 100%.

Bunge simply fails to realize that medical (as well as biological and social) observations are never “facts” because we deal with open and systems and hence uncertain inferences.

4.2. RCTs and the truth claim

We are now investigating Bunge’s claim that RCTs are necessary to infer the truth of a causal hypothesis in more detail. To this aim, it is helpful to first review some methodological principles of RCTs. A good overview has recently been provided by Deaton and Cartwright [27], and we follow their account to a large extent. Without loss of generality we assume that the medical hypothesis to be tested in a RCT is a proposition of the form “treatment T is effective” the truth of which is typically assessed by measuring some particular outcomes in the randomly allocated treatment and control groups. To measure the truth of a medical hypothesis then means to measure the true average

treatment effect (ATE) of the intervention, where the ATE is the difference between the average outcome in the treatment group and the average outcome in the control group.¹³ Assuming a linear causal model for the individual treatment effects one could write for an individual outcome

$$Y_i = \beta_i T_i + \sum_{j=1}^J \gamma_j x_{ij}.$$

Here Y_i is the outcome for patient i , T_i is a treatment indicator ($T_i = 1$ if treatment, $T_i = 0$ if control), β_i the individual treatment effect for patient i , and the x 's are observed or unobserved other linear causes of the outcome. By averaging the effects in both treatment (T) and control (C) group and subtracting the means one obtains an estimate for the ATE

$$\bar{Y}_T - \bar{Y}_C = \bar{\beta}_T + \sum_{j=1}^J \gamma_j (\bar{x}_{Tij} - \bar{x}_{Cij}) \quad (\text{ATE})$$

The major interest in conducting a RCT is on $\bar{\beta}_T$ which is the true ATE in case that the averages of the other causes are exactly balanced between both groups. Bunge claims that the aim of randomization is to bring the error term on the right-hand side of the (ATE) equation as close to zero as possible. However, this is not what any RCT can guarantee [24,26]. What randomization actually does is guaranteeing that the error term is zero only *in expectation*. The expectation refers to an infinite number of repeated randomizations of the trial sample into treatment and control group – for an individual randomization the estimated ATE can be arbitrarily far away from the true ATE. Repeating the trial and estimating the ATE many times allows one to estimate a mean ATE and its standard error – this is the true benefit of randomization. Contrary to what Bunge claims, therefore, randomization will not guarantee that an individual RCT will provide us with an estimate of the ATE that is close to the truth. Instead, if there is background knowledge about the main other causes of the outcome one would be better off matching patients according to these other causes without randomization. But background knowledge is exactly what is omitted by Bunge when he proposes RCTs as the gold standard of clinical trials. Deaton and Cartwright put it this way:

The gold standard or “truth” view does harm when it undermines the obligation of science to reconcile RCTs results with other evidence in a process of cumulative understanding.

The conception of patients as open systems forces us to accept that we can never infer the *true* effect of a treatment through RCTs, even if we would be able to repeat one and the same trial an infinite number of times. The reason is that in each repetition, some changes in the environment or context in which the RCT is conducted are unavoidable. The best we can therefore do is to seek higher and higher confirmation for our hypotheses and determine their evidence against realistic competing hypotheses. But these goals are achievable by collecting relevant data across a variety of study types of both statistical and mechanistic character. A famous example is the establishment of the hypothesis that smoking causes lung cancer that was based on observational and laboratory data, but not on RCTs. Surprisingly, Bunge himself has used this example in one of his previous papers [16]:

For example, since the mid-20th century, it has been known that lung cancer and smoking are strongly correlated, but only laboratory experiments on the action of nicotine and tar on living tissue have succeeded in testing (and confirming) the hypothesis that there is a definite causal link underneath the statistical correlation: we now know definitely that smoking may cause lung cancer.

Note that for Bunge it was knowledge of the mechanisms that confirmed the hypothesis that smoking causes lung cancer. However, on our account, the strong correlational data between smoking

¹³ Actually, the very significance of knowing the average treatment effect for medical practice may be questioned. See Subramanian et al. [56] for a brief commentary.

and lung cancer on its own provided a strong degree of confirmation for the hypothesis that smoking causes lung cancer. Knowing the mechanism regarding how smoking may cause lung cancer provided an additional, independent confirmation of this hypothesis so that the total confirmation became higher than with either the statistical or mechanistic data alone.¹⁴ However, causation was not established by the observational studies alone since the data they provided for a direct causal relationship between smoking and lung cancer was interpreted as not providing strong enough evidence compared to alternatives such as a “smoking gene” increasing both the tendency to smoke and to develop lung cancer.¹⁵ Only by knowing the carcinogenic mechanisms of tar and nicotine directly linking smoking and lung tumorigenesis was the observed correlation to be interpreted as strong data that smoking directly causes lung cancer instead of both being due to some third factor. Given our likelihood-based account of evidence, which must be comparative, we have three possible ways to compare two hypotheses. One could be a comparison between a causal hypothesis and a statistical hypothesis. The second could be a comparison between a causal hypothesis with another causal hypothesis. The third and final one could be between a statistical and another statistical hypothesis. The current scenario is concerned with case one in which a comparison is made between a causal hypothesis and a non-causal statistical hypothesis. Given the accumulated data regarding several observational studies on the proportion of tar in tobacco, the hypothesis that smoking causes cancer was supported more strongly than the hypothesis that smoking and cancer were merely correlated without implying any causal connection between them. This evidential relationship between the hypotheses, smoking causes cancer, and smoking and cancer are statistically related, given the data, holds independent of what an agent believes about those hypotheses and data. Therefore, from the basis of our account of evidence, we could say that data provide strong evidential support for the hypothesis that smoking causes cancer as against its alternative. So, from the perspectives of both accounts of evidence and confirmation, the hypothesis “smoking causes cancer” is more evidentially supported by data than its alternative hypothesis as well as it is strongly confirmed. This both-way vindication is possible because of the theorem: If two hypotheses are mutually exclusive and jointly exhaustive as well as simple statistical hypotheses, then data will provide evidential support for a hypothesis over its alternative if and only if data will confirm the hypothesis to some degree:

$$[\text{Pr}(D|H)/\text{Pr}(D|\neg H)] > 1 \text{ iff } \text{Pr}(H|D) > \text{Pr}(H).$$

Another example, also mentioned by Bunge himself [1, p. 147], is appendectomy to treat appendicitis. In this case, conducting a RCT with a control group receiving sham operation would not only be unethical, but also not necessary to highly confirm the hypothesis that appendectomy is an effective treatment. The reason is that the mechanism (infection of the appendix) and the way to shut it down (by removing the appendix) are very well known. This example is noteworthy since it could serve to illustrate that a causal hypothesis may be established by knowledge of mechanisms alone without the necessity for statistical data.¹⁶ A third example is the treatment of the rare disease glucose transporter 1 (GLUT1) deficiency syndrome through prescribing a high-fat low carbohydrate ketogenic diet: despite only “low level” clinical evidence available, a recent consensus guideline recommends ketogenic diets as the treatment of choice for GLUT1 deficiency syndrome mainly based

¹⁴ This is a consequence of the so-called “variety-of-evidence thesis” [42,43]. On our account, this is a slight misnomer. We call this the “variety-of-data thesis”.

¹⁵ See chapter 5 in Judea Pearl’s “Book of Why” [57] for a detailed historical summary on how causation between smoking and lung cancer became established, including the argument that a “smoking gene” might be the underlying confounding factor.

¹⁶ The example is a counter-example to the Russo-Williamson thesis according to which both statistical and mechanistic evidence is required to establish a causal claim [58]. See also Claveau [59] for another counter-example to the Russo-Williamson thesis from the social sciences.

on the physiological mechanism that ketone bodies are able to cross the blood-brain barrier independent from GLUT1, providing an alternative fuel for the brain instead of glucose [39].¹⁷

As these examples show, RCTs are neither necessary nor sufficient for determining whether some factors cause a disease or an intervention is effective. Rather, causal claims may be established based on a variety of data from mechanistic, observational and other study types that conventionally sit below RCTs in the “evidence hierarchy”. We note that the difference between mechanistic and statistical (or probabilistic) data is not one between qualitative and quantitative data, nor one between observations stemming from laboratory versus clinical studies. In fact, mechanistic hypothesis may be framed as statistical models and applied to clinical data. For example, in radiotherapy mathematical models describing the mechanisms of cell killing through DNA damage caused by ionizing radiation are frequently utilized. They may be used clinically to convert between different fractionation schemes having the same biological effect or for predicting radiotherapy outcomes such as tumor control and normal tissue complication probability [40,41]. In our interpretation, the main distinction between mechanistic and statistical data is that the former can explain *why* the latter are observed. In the radiotherapy example, the mechanism itself is stochastic as it describes the killing of cells which obeys statistical laws; however, it also explains *why* higher radiation doses result in higher probability of tumor control, and even allows the derivation of the mathematical form of the dose-response relationship [40].

Each causal claim has both probabilistic and mechanistic consequences that may be observed or not. Therefore, either mechanistic or statistical data are able to confirm a causal claim to some degree, whereas data from both sources provide even stronger confirmation according to the “variety-of-evidence thesis” [42,43]. Mechanisms can also serve as background knowledge to increase the evidence of a causal hypothesis over a merely statistical one; this was the case in the smoking and lung cancer example. Finally, the optimal methodology for establishing a causal claim may depend on the exact type of hypothesis posed, e.g., the claim that an intervention worked in some setting versus the claim that it works for a particular patients, or a claim about a harmful effect [28].

4.3. RCTs and background knowledge

For any researcher, prior or background knowledge plays a crucial role for the evaluation of causal hypotheses. This is naturally captured in our account of evidence and confirmation described in section 4.1, but not in Bunge’s account relying on RCTs as necessary methods for hypothesis testing. Andrew Gelman [44] has pointed out that using tools such as randomization and p-values to enforce scientific rigor misses the most important point of causal inferences, which is interpreting and understanding the results within the context of background knowledge. RCTs are not designed to rely on background knowledge which “is an advantage when persuading distrustful audiences, but it is a disadvantage for cumulative scientific progress, where prior knowledge should be built upon, not discarded” [27]. Thus, demanding the conduct of a RCT as a necessary condition for confirming a causal hypothesis, as Bunge does, violates his own postulates (S4) and (S5), because it discourages grouping hypotheses into medical theories and makes only minimal assumptions about the validity of other benchmark items, i.e., background knowledge. Judea Pearl has emphasized that if an agent is able to use her background knowledge in order to frame a causal model of reality, RCTs are no longer the only means to estimate the effect of interventions. In this case observational studies can do just as good with two additional advantages: their conduction is often more practical, and they study populations in their natural environment, instead of an artificial environments created by experimental

¹⁷ Ketone bodies are produced under conditions of low insulin levels such as during fasting or very low carbohydrate intake; in this respect, ketogenic diets can mimic fasting without necessarily restricting energy intake. Ketone bodies are transported through the blood-brain barrier by monocarboxylate transporters; this transport mechanism is therefore completely insulin- and GLUT-independent.

protocols [45]. It is also well known that RCTs are not immune to bias, so that poorly designed RCTs may provide less certain results than well designed observational studies [27]. Background knowledge of the structure and mechanisms in the system under study is also important for meaningfully interpreting RCT results, or generally results from any statistical study. Deaton and Cartwright illustrate this using Bertrand Russel's famous chicken example [27, p.11]:

The bird infers, on repeated evidence, that when the farmer comes in the morning, he feeds her. The inference serves her well until Christmas morning, when he wrings her neck and serves her for dinner. Though this chicken did not base her inference on an RCT, had we constructed one for her, we would have obtained the same result that she did. Her problem was not her methodology, but rather that she did not understand the social and economic structure that gave rise to the causal relations that she observed.

The importance of taking background knowledge into account also arises each time that results are discordant either between individual RCTs or between a RCT and another study type. From a purely empiricist standpoint ignoring what we know about the interplay between an intervention, the context under which it is applied and its mechanisms, such discrepancies are usually explained by invoking certain quality criteria based on design and statistical arguments [46]. For example, if there is discrepancy between RCTs and observational studies, results from the former are usually taken to "override" results from the latter. However, acknowledging that observing an intervention effect presupposes some mechanisms at work whose activation in turn may be context-dependent opens up much more possibilities for interpreting negative study results or discrepancies between study types. In particular, studies investigating an intervention may vary in context, in the mechanism that is exerted, or in both simultaneously [46]. An example for the latter situation is the supplementation of antioxidant vitamins for preventing cardiovascular disease which have been declared ineffective based on mostly negative findings in RCTs, although evidence from observational studies showed preventive effects. Connelly [46] emphasizes the realist standpoint that different antioxidant vitamins may act via different mechanisms that in turn may depend on the age of an individual and whose effects may only be observed over much longer follow-up periods than usually used in RCTs. He concludes:

It seems that an alternative realistic perspective on this question [whether antioxidant vitamins can prevent cardiovascular disease] is again ignored in favour of what purports to be an unassailable scientific observation of the results from RCTs. Here, once more, the effect of ignoring differences in mechanisms and contexts may be to close down research in this area prematurely.

By claiming that only RCTs can establish or not establish the efficacy of medical interventions, both Bunge and EBM discourage realist thinking about mechanisms and contexts whenever RCT results are available. As the antioxidant example shows, such thinking may preclude scientific progress, especially when interventions and effects are related via complex mechanisms. This is despite EBM explicitly stating the importance of evaluating "the totality of evidence" as one of its epistemological principles [19]. Maybe this is also the reason why Bunge downgrades complementary and alternative medicine (CAM) as "unscientific", because by its very nature CAM works with complex interventions differently from simple drug administrations for which an evidence hierarchy with RCTs on top appears inadequate [29]. While for Bunge mechanisms are essential for understanding empirical phenomena through what he calls *mechanismic* explanation [16], he restricts their main epistemological role within the context of scientific medicine to "boosting" the confirmation of causal hypotheses provided by RCTs – in fact, no mention is made in his book how mechanisms may be used in conjunction with methodologies other than RCTs. In our opinion, this underestimates the role mechanisms should play in medical hypothesis testing and treatment design. As the examples of appendectomy and GLUT1 deficiency syndrome given in subsection 4.2 show

there are situations where data of mechanisms become equally or more important than statistical data for establishing a causal claim of treatment efficacy.

As another example, consider the establishment of a causal relationship between benzo[a]pyrene exposure and carcinogenesis in humans despite a lack of clinical data. Wilde and Parkkinen [47] have argued that one is justified in believing that benzo[a]pyrene causes cancer in humans because animal studies have provided evidence for both the robustness of the causal association and the mechanisms at work. In this example, knowledge of mechanisms provides the basis for the extrapolation of study results across different contexts, or more generally for building a causal theory that can then be applied to varying contexts. Such a theory resting on *mechanismic* explanation is much more flexible than individual hypotheses. In particular, predictions can be made from one context to another. At the same time, once particular mechanisms have confirmed a causal theory, “we should attempt to eliminate alternative explanations by testing the potential effects of these mechanisms, particularly in contexts other than the one where the theory was created” [48]. In other words, we should try to establish evidence that the mechanisms that are part of our theory are at work across a variety of contexts. To this aim, even case studies are valuable tools because they usually study individuals within more natural environments different from the ones artificially created by clinical study protocols and at the same time perform more thorough measurements than epidemiological studies.

5. Conclusions

Mario Bunge has provided a unique treatise of the philosophy of medicine which is based on a realist-systemic ontology. In Bunge’s view, the world consists of systems, which in turn consist of components, structures, mechanisms and an environment. These four constituents should be taken into account when studying any individual system. Bunge supports the methodology of EBM which considers RCTs to possess the most rigorous study design and to provide the best evidence for establishing the truth of causal hypotheses. Both Bunge and EBM thereby adopt a Humean view of causality in which causes and effects are constant conjunctions of events which in Bunge’s ontology are mediated through mechanisms. However, as we have argued in this paper, the view of RCTs as the best or even the only methodology for achieving knowledge about the efficacy of medical interventions is incompatible with a realist-systemic ontology. The reason is that the Humean view of causality only holds in closed systems, but patients and study populations are heterogenous open systems, and even the most rigorously conducted RCTs cannot guarantee closure of the system under study. This implies that the Humean view of causality cannot be sustained in medicine. Hence, Bunge’s claim that RCTs are necessary to establish the truth of the efficacy of an intervention must be rejected.

Another consequence of the failure of Bunge’s and EBM’s Humean view of causality is that we need a probabilistic account of both evidence and confirmation. We have provided such accounts in this paper, building on the subjective Bayesian concept of confirmation and the objective concept of a likelihood ratio or Bayes factor to compare two competing hypotheses. Our Bayesian account of evidence is compatible with a realist-systemic ontology insofar as evidence is conceived as an objective function of the data generating process, which takes place via observed or unobserved mechanisms within the system under study. Finally, we maintain that Bunge’s general critique of Bayesianism is based on wrong interpretations of central concepts of probability, as the AIDS example discussed in section 4.1 has shown.

We agree with Bunge that statistical data alone are not sufficient for causal reasoning; instead, what is needed is a causal hypothesis or causal model which can only be arrived at by thinking about structures and mechanisms. However, Bunge must accept the critique of overrating the merits of RCTs since he fails to account for the epistemological role of both already established mechanisms and statistical associations observed in different contexts. In fact, the environment which determines

whether causal effects will emerge or not, is almost completely neglected in his account of medical hypothesis testing. His conception of an evidence hierarchy is at best unnecessary, at worst however an obstacle to scientific progress in the design and evaluation of medical treatments for individual patients. An analogous proposition applies to EBM with its various forms of methodological hierarchies [28]. The role of background knowledge of both mechanisms and the stability of statistical regularities across various contexts must be taken into account, which requires a variety of research methodologies. For example, critical realism, which has an ontology similar to Bunge's systemism¹⁸, embraces a multitude of methodologies co-existing along each other, with no particular general preference for one methodology over any other [49,50]. We, along with an increasing number of critical philosophers and medical methodologists [30], encourage a similar approach in medicine, where "[t]he important point is not whether a study is randomized or not, but whether it uses a method well suited to answer a question and whether it implements this method with optimal scientific rigor" [29, p.5]. This implies considering results from preclinical and theoretical modelling studies, non-randomized cohort studies, observational and case studies along with those from RCTs (if available) in an integrated or "circular" [29] rather than dogmatic-hierarchical framework, evaluating evidence based on those methodologies that best provide the data relevant to the particular hypothesis under investigation.

Acknowledgements

We are thankful to Gordan G. Brittan Jr. for providing valuable comments on an earlier version of this manuscript.

Author contributions

Conceptualization, RJK; Methodology, RJK and PSB; Writing – Original Draft Preparation, RJK.; Writing – Review & Editing, RJK and PSB; Visualization, RJK; Supervision, PSB;

Conflicts of Interest

The authors declare no conflict of interest.

REFERENCES

1. Bunge, M. *Medical Philosophy: Conceptual Issues In Medicine*; 1st ed.; World Scientific Publishing Co Pte Ltd: Singapore, 2013; ISBN 9814508942.
2. Howick, J. *The Philosophy of Evidence-Based Medicine*; 1st ed.; John Wiley & Sons Ltd: Oxford, UK, 2011; ISBN 978-1-4051-9667-3.
3. Thompson, R.P.; Upshur, R.E.G. *Philosophy of Medicine*; 1st ed.; Routledge: New York, 2017; ISBN 978-0-415-50109-5.
4. Parkkinen, V.P.; Wallmann, C.; Wilde, M.; Clarke, B.; Illari, P.; Kelly, M.P.; Norell, C.; Russo, F.; Shaw, B.; Williamson, J. *Evaluating Evidence of Mechanisms in Medicine: Principles and Procedures*; 1st ed.; Springer International Publishing AG: Cham, Switzerland, 2018; ISBN 978-3-319-94609-2.
5. Gillies, D. *Causality, Probability, and Medicine*; 1st ed.; Routledge: New York, 2019; ISBN 978-1-138-82930-5.
6. Ahn, A.C.; Tewari, M.; Poon, C.-S.; Phillips, R.S. The Limits of Reductionism in Medicine: Could Systems Biology Offer an Alternative? *PLoS Med.* **2006**, *3*, e208.
7. Ahn, A.C.; Tewari, M.; Poon, C.-S.; Phillips, R.S. The Clinical Applications of a Systems Approach. *PLoS Med.* **2006**, *3*, e209.

¹⁸ Critical realism maintains the naïve realistic view of an independently existing world of objects and structures giving rise to events that do and do not occur, while at the same time acknowledging the epistemological limitations of our observations and knowledge that are relative to our time period and culture [50]. Emergence and the concept of systems are central themes in critical realism [50], similar to Bunge's systemism.

8. Kesić, S. Systems biology, emergence and antireductionism. *Saudi J. Biol. Sci.* **2016**, *23*, 584–591.
9. Walach, H.; Loughlin, M. Patients and agents – or why we need a different narrative: A philosophical analysis. *Philos. Ethics, Humanit. Med.* **2018**, *13*, 13.
10. Schoenfeld, J.D.; Ioannidis, J.P.A. Is everything we eat associated with cancer? A systematic cookbook review. *Am J Clin Nutr* **2013**, *97*, 127–134.
11. Cofnas, N. Methodological problems with the test of the Paleo diet by Lamont et al. (2016). *Nutr. Diabetes* **2016**, *6*, e214.
12. Bains, W.; Schulze-Makuch, D. The Cosmic Zoo: The (Near) Inevitability of the Evolution of Complex, Macroscopic Life. *Life* **2016**, *6*, 25.
13. Bunge, M. Emergence and the Mind. *Neuroscience* **1977**, *2*, 501–509.
14. von Bertalanffy, L. An outline of general system theory. *Br. J. Philos. Sci.* **1950**, *1*, 134–165.
15. Bunge, M. Systemism: The alternative to individualism and holism. *J. Socio. Econ.* **2000**, *29*, 147–157.
16. Bunge, M. Mechanism and Explanation. *Philos. Soc. Sci.* **1997**, *27*, 410–465.
17. Bunge, M. How does it work? The search for explanatory mechanisms. *Philos. Soc. Sci.* **2004**, *34*, 182–210.
18. von Bertalanffy, L. Basic concepts in quantitative biology of metabolism. *Helgoländer Wissenschaftliche Meeresuntersuchungen* **1964**, *9*, 5–37.
19. Djulbegovic, B.; Guyatt, G.H. Progress in evidence-based medicine: a quarter century on. *Lancet* **2017**, *390*, 415–423.
20. Miles, A.; Bentley, P.; Polychronis, A.; Grey, J. Evidence-based medicine: why all the fuss? This is why. *J. Eval. Clin. Pract.* **1997**, *3*, 83–86.
21. Welsby, P.D. Reductionism in medicine: Some thoughts on medical education from the clinical front line. *J. Eval. Clin. Pract.* **1999**, *5*, 125–131.
22. Sniderman, A.D.; LaChapelle, K.J.; Rachon, N.A.; Furberg, C.D. The necessity for clinical reasoning in the era of evidence-based medicine. *Mayo Clin. Proc.* **2013**, *88*, 1108–1114.
23. Klement, R.J.; Bandyopadhyay, P.S.; Champ, C.E.; Walach, H. Application of Bayesian evidence synthesis to modelling the effect of ketogenic therapy on survival of high grade glioma patients. *Theor. Biol. Med. Model.* **2018**, *15*, 12.
24. Worrall, J. What Evidence in Evidence-Based Medicine? *Philos. Sci.* **2002**, *69*, S316–S330.
25. Goldenberg, M.J. On evidence and evidence-based medicine: Lessons from the philosophy of science. *Soc. Sci. Med.* **2006**, *62*, 2621–2632.
26. Urbach, P. The value of randomization and control in clinical trials. *Stat. Med.* **1993**, *12*, 1421–1431.
27. Deaton, A.; Cartwright, N. Understanding and misunderstanding randomized controlled trials. *Soc. Sci. Med.* **2018**, *210*, 2–21.
28. Stegenga, J. Down with the Hierarchies. *Topoi* **2014**, *33*, 313–322.
29. Walach, H.; Falkenberg, T.; Fønnebø, V.; Lewith, G.; Jonas, W.B. Circular instead of hierarchical: methodological principles for the evaluation of complex interventions. *BMC Med. Res. Methodol.* **2006**, *6*, 29.
30. Anjum, R.L.; Copeland, S.; Rocca, E. Medical scientists and philosophers worldwide appeal to EBM to expand the notion of ‘evidence.’ *BMJ Evidence-Based Med.* **2018**.
31. Clarke, B.; Gillies, D.; Illari, P.; Russo, F.; Williamson, J. Mechanisms and the Evidence Hierarchy. *Topoi* **2014**, *33*, 339–360.
32. Lawson, T. Abstraction, tendencies and stylised facts: a realist approach to economic analysis. *Cambridge J. Econ.* **1989**, *13*, 59–78.
33. Brannan, M.J.; Fleetwood, S.; O’Mahoney, J.; Vincent, S. Critical Essay: Meta-analysis: A critical realist critique and alternative. *Hum. Relations* **2017**, *70*, 11–39.
34. Bandyopadhyay, P.S.; Brittan, G.G. Acceptability, evidence, and severity. *Synthese* **2006**, *148*, 259–293.
35. Bandyopadhyay, P.S.; Brittan Jr., G.; Taper, M.L. *Belief, Evidence, and Uncertainty: Problems of Epistemic Inference*; 1st ed.; Springer International Publishing: Basel, 2016; ISBN 978-3-

- 319-27770-7.
36. Burnham, K.P.; Anderson, D.R. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociol. Methods Res.* **2004**, *33*, 261–304.
 37. Kass, R.E.; Raftery, A.E. Bayes Factors. *J. Am. Stat. Assoc.* **1995**, *90*, 773–795.
 38. Anzala, O.; Sanders, E.J.; Kamali, A.; Katende, M.; Mutua, G.N.; Ruzagira, E.; Stevens, G.; Simek, M.; Price, M. Sensitivity and specificity of HIV rapid tests tests used for research and voluntary conunselling and testing. *East Afr. Med. J.* **2008**, *85*, 500–504.
 39. Kossoff, E.H.; Zupec-Kania, B.A. Optimal clinical management of children receiving the ketogenic diet: recommendations of the International Ketogenic Diet Study Group. *Epilepsia Open* **2018**, *3*, 175–192.
 40. Fowler, J.F. 21 Years of biologically effective dose. *Br. J. Radiol.* **2010**, *83*, 554–568.
 41. Klement, R.J.; Allgäuer, M.; Andratschke, N.; Blanck, O.; Boda-Heggemann, J.; Dieckmann, K.; Duma, M.; Ernst, I.; Flentje, M.; Ganswindt, U.; et al. Bayesian Cure Rate Modeling of Local Tumor Control: Evaluation in Stereotactic Body Radiotherapy for Pulmonary Metastases. *Int. J. Radiat. Oncol.* **2016**, *94*, 841–849.
 42. Bovens, L.; Hartmann, S. *Bayesian Epistemology*; 1st ed.; Oxford University Press: New York, 2003; ISBN 978-0199270408.
 43. Claveau, F.; Grenier, O. The variety-of-evidence thesis: a Bayesian exploration of its surprising failures. *Synthese* **2017**, 1–28.
 44. Gelman, A. Benefits and limitations of randomized controlled trials: A commentary on Deaton and Cartwright. *Soc. Sci. Med.* **2018**, *210*, 48–49.
 45. Pearl, J. Challenging the hegemony of randomized controlled trials: A commentary on Deaton and Cartwright. *Soc. Sci. Med.* **2018**, *210*, 60–62.
 46. Connelly, J. Realism in evidence based medicine: interpreting the randomised controlled trial. *J. Health Organ. Manag.* **2004**, *18*, 70–81.
 47. Wilde, M.; Parkkinen, V.P. Extrapolation and the Russo–Williamson thesis. *Synthese* **2017**, 1–12.
 48. Tsang, E.W.K. Case studies and generalization in information systems research: A critical realist perspective. *J. Strateg. Inf. Syst.* **2014**, *23*, 174–186.
 49. Mingers, J. A critique of statistical modelling in management science from a critical realist perspective: Its role within multimethodology. *J. Oper. Res. Soc.* **2006**, *57*, 202–219.
 50. Mingers, J. *Systems Thinking, Critical Realism and Philosophy: A Confluence of Ideas*; 1st ed.; Routledge: New York, 2014; ISBN 978-0415519533.
 51. Aronson, J.L. *A Realist Philosophy of Science*; 1st ed.; The Macmillan Press Ltd: London, 1984; ISBN 978-1-349-17380-8.
 52. Iftikhar, H.; Saleem, M.; Kaji, A. Metformin-associated Severe Lactic Acidosis in the Setting of Acute Kidney Injury. *Cureus* **2019**, *11*, e3897.
 53. MacKay, D.J.C. *Information Theory, Inference, and Learning Algorithms*; 3rd ed.; Cambridge University Press: Cambridge, United Kingdom, 2004;
 54. Bailer-Jones, C.A.L. *Practical Bayesian Inference. A Primer for Physical Scientists.*; Cambridge University Press: Cambridge, United Kingdom, 2017; ISBN 978-1-316-64221-4.
 55. Klement, R.J. Beneficial effects of ketogenic diets for cancer patients: a realist review with focus on evidence and confirmation. *Med. Oncol.* **2017**, *34*, 132.
 56. Subramanian, S. V.; Kim, R.; Christakis, N.A. The “average” treatment effect: A construct ripe for retirement. A commentary on Deaton and Cartwright. *Soc. Sci. Med.* **2018**, *210*, 77–82.
 57. Pearl, J.; Mackenzie, D. *The Book of Why: The New Science of Cause and Effect*; 1st ed.; Basic Books: New York, 2018; ISBN 978-0465097609.
 58. Russo, F.; Williamson, J. Interpreting causality in the health sciences. *Int. Stud. Philos. Sci.* **2007**, *21*, 157–170.
 59. Claveau, F. The Russo-Williamson Theses in the social sciences: Causal inference drawing on two types of evidence. *Stud. Hist. Philos. Sci. Part C Stud. Hist. Philos. Biol. Biomed. Sci.* **2012**, *43*, 806–813.