UDK: 524.6:14
DOI: 10.2298/FID1304143C
Original scientific paper

Milan M. Ćirković
Astronomical Observatory of Belgrade
Future of Humanity Institute
Oxford University

Counterfactuals and unphysical ceteris paribus: an explanatory fallacy

Abstract *I reconsider a type of counterfactual argument often used in historical sciences on a recent widely discussed example of the so-called "rare Earth" hypothesis in planetary sciences and astrobiology. The argument is based on the alleged "rarity" of some crucial ingredient for the planetary habitability, which is, in Earth's case, provided by contingent evolutionary development. For instance, the claim that a contingent fact of history which has created planet Jupiter enables shielding of Earth from most dangerous impact catastrophes, thus increasing Earth's habitability, leads often to the conclusion that such state-of-affairs must be rare in the Galaxy. I argue that this reasoning is deeply flawed, for several closely related reasons. In addition, the relevance of the philosophical problem of transworld identity for this kind of historical reasoning in science is put forward. This highlights many explanatory problems one faces when using historical counterfactuals in study of complex, nonlinear dynamical systems – and bolsters the relevance of philosophy for evaluation of scientific explanatory claims.*

Keywords: *counterfactuals, astrobiology, philosophy of science, philosophy of history, transworld identity*

Introduction: the rare earth hypothesis

Recent years showed a tremendous increase of interest in explanatory strategies based on counterfactual reasoning in many areas of science and philosophy. Most notably, the emergence of "virtual history" (Ferguson 1999) in historical sciences, and somewhat contemporary reappraisal of the role of contingency in natural sciences, notably planetary sciences, origin of life (abiogenesis) studies and evolutionary biology. In the latter context, one of the most controversial issues arose in the nascent discipline of *astrobiology*, which deals with three canonical questions: How does life begin and develop? Does life exist elsewhere in the universe? What is the future of life on Earth and in space? (For reviews see Des Marais and Walter 1999; Darling 2001; Grinspoon 2003; Chyba and Hand 2005; Blumberg 2011.) However, the epistemological and methodological basis of astrobiological studies presents us with a hornet's nest of issues which have not been, with several exceptions,

tackled in the literature so far (Ćirković 2012). It is not surprising, therefore, that seemingly paradoxical situations and conclusions arise from time to time, as is usual in young scientific fields – some of which might be instructive from a very general point of view.

The "rare-Earth hypothesis" (henceforth REH), put forward by Peter Ward and Donald Brownlee in 2000, has attracted a lot of attention as perhaps the prototypical "astrobiological theory" bold enough to tackle issues on the Galactic – and larger – scales. In a nutshell, REH is a probabilistic argument suggesting that, while simple microbial life is probably ubiquitous throughout the Galaxy, complex biospheres, like the terrestrial one, are very rare due to the exceptional combination of many distinct requirements. These ingredients of the REH are well known to even a casual student of astrobiology:

- **Galactic Habitable Zone:** A habitable planetary system needs to be in a narrow annular ring in the Milky Way disc, where chemical abundances and stability conditions for the emergence and evolution of life are satisfied.
- **Circumstellar habitable zone:** A habitable planet needs to be in the very narrow interval of distances from the parent star in order to possess liquid water on surface.
- **"Rare Moon":** Having a large moon to stabilize the planetary axis is crucial for long-term climate stability.
- **"Rare Jupiter":** Having a giant planet ('Jupiter') at the right distance to deflect much of incoming cometary and asteroidal material enables a sufficiently low level of impact catastrophes.
- **"Rare nuclides":** Radioactive *r*-elements (especially ^{238}U and ^{232}Th) need to be present in the planetary interior in sufficient amounts to enable plate tectonics and the functioning of the carbon–silicate cycle.
- **"Rare Cambrian-explosion analogues":** The evolution of complex metazoans requires exceptional physical, chemical and geological conditions for episodes of sudden diversification and expansion of life.

There are other items on the list as well – in this sense, REH is an open theoretical system, since everyone is free to add items pertaining to a particular area of relevance and expertise. However, the general reasoning is that all these requirements are mostly independent and *a priori*

unlikely, so that their combination is bound to be incredibly rare and probably unique in the Milky Way (Ward and Brownlee 2000, p. 283):

> The continued marginalization of Earth and its place in the Universe perhaps should be reassessed. We are not the center of the Universe, and we never will be. But we are not so ordinary as Western science has made us to be for two millennia. Our global inferiority complex may be unwarranted. What if Earth is extremely rare because of its animals (or, to put it another way, because of its animal habitability)?

Ward and Brownlee even construct a symbolic equation they dub the 'Rare Earth equation' – by analogy with the Drake equation often (mis) used in SETI studies – which contains the requirements as fractions of the total set of Galactic planets. The product of many small probabilities is a much smaller probability, so there are few places, if any, satisfying all the REH criteria. The fact that we observe such a place around us should not be surprising, since observation selection tells us that we could not observe anything else; *there are no observers in places where the requirements are not met*.

Ward and Brownlee make it very clear that they do not regard the transition between non-living and living matter as particularly difficult. In this regard, they follow closely the *continuity thesis* of Iris Fry, (1995, 2000): there is no big gap between non-living and living and abiogenesis will happen rather quickly where physical, chemical, geological, etc. conditions are satisfied. I shall briefly return to this important philosophical issue in the concluding section.

It is complex life, on Earth following the invention of multicellularity and, especially, the dramatic diversification in the Cambrian explosion, which is cosmologically rare. Therefore, it appears that REH is considerably easy to falsify by either finding an absence of simple life forms or finding traces of complex life forms (including intelligent ones; more on the operational definition of intelligence later). In particular, a clear prediction of REH is that we neither should be surprised to find living or extinct microbial life forms on Mars, nor biomarkers such as free atmospheric oxygen on Earth-like extrasolar planets; another prediction is that SETI projects, embodying searches for the most complex known life forms, will remain unsuccessful.

In addition, Ward and Brownlee break new ground by pointing out the importance of hitherto downplayed factors, like plate tectonics, inertial interchange events, or 'Snowball Earth' episodes of global glaciation, for

the development of complex life. This wide scope was perhaps the key factor why REH left such a strong impression in the decade following publication of their book, and the reason why it quickly became somewhat of a default position in many astrobiological circles. Thus, its challenge to Copernicanism has been partially accepted as sound in mainstream astrobiology – although, as argued by opponents, and as I have argued elsewhere (Ćirković 2012), there are lower prices to be paid in the market of ideas. Even very strong criticisms of REH just underline its importance in exemplifying astrobiological theory and highlighting specific predictions. Something very similar occurred during the time of the 'Great Controversy' in cosmology: the steady state theory had been – although wrong – a crucial trigger of the great progress in understanding the large-scale structure of spacetime. Can REH play such a role, mutatis mutandis, in the astrobiological realm? As far as the reasoning of Ward and Brownlee goes, one can hardly call it into question. As seen in Impey's (2010) book, just a minority of practising astrobiologists openly subscribe to REH. How is scepticism of the rest justified? There are essentially two lines of attack: (i) to deny the independence of the various particular requirements (items listed above), and (ii) to deny that the particular requirements are indeed unlikely. In addition to these two, there is a vaguer sort of argument to the effect that the terrestrial biosphere might not be representative of the entire class of 'complex biospheres', so that the requirements for the terrestrial subtype might be more restrictive than in the general case.

However, a very important general argument, mostly downplayed in astrobiology thus far, is hidden within type (i) criticisms of REH denying independence of the requirements and questioning the *ceteris paribus* part of the argument. Its philosophical implications are multifold, and therefore a more detailed discussion in the rest of this paper is warranted. Some it is almost verbatim applicable to other fields containing complex historical systems (cosmology, paleontology, evolutionary biology, political and military history, etc.).

Rare Earth hypothesis and counterfactuals

Suppose somebody claimed that the twin paradox of special relativity invalidates that theory, since both twins cannot be each younger than the other. Upon demonstrating the asymmetry of twins' positions due to acceleration of spaceship at turn-about, a relativity skeptic – if hard-headed and a bit irrational – could only continue to insist that there can be no

such thing, since in the thought-experiment setup *there was no mention of acceleration*. In other words, she ends up falling into a huge 'coherence gap' – in the words of the contemporary philosopher Ivan Havel (1999):

> In conceivable worlds of thought experiments, some states-of-affairs are, by design, the same as they are in our world, while other states-of-affairs are deliberately different... The crucial but often neglected feature of these worlds is that we seldom know what is the extent of the domain of "the same" and what is the extent of the domain of "the different", besides what is explicitly mentioned or used in the construction. Moreover, besides these two domains there is an inexhaustible realm of states-of-affairs that are omitted because they are believed to be irrelevant or because they are forgotten, obscured or entirely beyond the reach of human knowledge.

The omitted realm of states of affairs constitutes the coherence gap, and the question of its possible impact on our reasoning, the coherence gap problem. We need to be highly cautious in evaluating any thought experiment not because it may entail empirical difficulties, but because the conceived world of the particular thought experiment may need additional assumptions or constraints in order for the desired outcome to occur.

Further examples of the same tendency, slightly less amusing, are provided by bad thought experiments in 'virtual history' (or bad historical counterfactuals). In an example suggested by Niall Ferguson, 'no sensible person wishes to know what would have happened in 1848 if the entire population of Paris had suddenly sprouted wings, as this is not a plausible scenario.' (Ferguson 1999, p. 83) But the plausibility is, at least, in part, in the eye of the beholder; while the implausibility of winged revolutionary Parisiennes seems obvious even to the layperson, there are certainly many cases of equal implausibility in 'exotic' fields of research – like astrobiology – which are not that obvious. Sometimes, they are extremely well hidden. Thus, the most prominent proponent of the use of counterfactuals to explain historical causation warns:[1]

> There are, in other words, two distinct kinds of counterfactual which have been used by historians: those which are essentially the products of imagination but (generally) lack an empirical basis; and those designed to test hypotheses by (supposedly) empirical means, which eschew imagination in favor of computation. In the case of the former, it is the tendency to rely for inspiration on hindsight, or to posit reductive explanations, which leads to implausibility. In the case of the latter, it is the tendency to make anachronistic assumptions.

1 Ferguson (1999), p. 18.

There is a huge literature on this topic, mainly belonging to philosophy of history and historiography.[2] It warns a careful thinker that the use of counterfactuals is a delicate matter and should be approached with extreme care, in spite of their popular appeal. It is exactly these 'anachronistic assumptions' which, as we shall see, constitute a major part of some of the REH requirements. In these examples, the coherence gap undermines reasoning by hiding additional assumptions of contingent or empirical nature (as contrasted with logical). In the context of historical sciences dealing with the material world – such as palaeontology, cosmology or cosmogony of the Solar System – I shall call this problem the unphysical *ceteris paribus*. It is a particular case of the wider fallacy of neglecting the context. If we are certain that A and B entail X (while X does not entail A and B), but for some reason we can observe or take into account only A – because of observational limitations, theoretical incompleteness or computational intractability – then it is often the case that observation of some other A' ≠ A would lead us to believe that ¬X. But the reasoning is incorrect, since it might be the case that some other B' holds, and A' and B' also entail X. In this simplified model, B is the *ceteris paribus*, which we wish to retain in order for our counterfactual thought experiment to be meaningful. However, it might be that the coherence gap makes the state of affairs (A',B) incoherent.

In the context of REH, unphysical *ceteris paribus* is visible in many places, since many of the individual Rare-Earth requirements are formulated in terms of Earth-related counterfactuals: *if such and such were different, the Earth would not be hospitable to complex life forms*. Thus, since such and such is *prima facie* improbable, this decreases the overall probability of complex life forms evolving anywhere. I submit that most of these arguments are unsound, since the implicit *ceteris paribus* cannot be maintained, which is visible in the well-known instance of "rare Jupiter".

The issue at hand is the famous argument about Jupiter being the optimal 'shield' of Earth from cometary/asteroidal bombardment. A smaller part of the problem is that the claim might be empirically wrong. The common-sense conclusion about the role of Jupiter – employed by REH – has been brought into question by the recent work of Horner and

2 Some of the entry points, apart from Ferguson's excellent anthology, are Hawthorne (1991) and Cowley (1999). A great precursor is "the three lives" essay(s) of Arnold J. Toynbee (1969). An amusing popular book by Comins (2010) shows how counterfactual approach might work in astronomy (though it suffers from many oversimplifications).

Jones (2008, 2009), who use massive numerical simulations to show that the conjecture that Jupiter acts as a shield against bombardment of the inner Solar System is untrue in a large part of parameter space. Moreover, they conclude, 'that such planets often actually increase the impact flux greatly over that which would be expected were a giant planet not present.' In other words, Jupiter might have a detrimental effect on the habitability of Earth! If the results of Horner and Jones withstand the test of time and further research, it is a serious blow to REH and the related way of thinking. However, it would still be a major score for astrobiology as a field, since it will demonstrate maturity of the discipline in which intuitively solid prejudices could be rejected based on precise, quantitative work.

However, the central issue of philosophical interest is that the *ceteris paribus* state of a "Jupiter-less" Solar System is unphysical. REH theorists argue that:

1) Both Earth and Jupiter exist in the Solar System.
2) Jupiter deflects a fraction of potential impactors from collision trajectories.
3) With more impactors on collision trajectories, Earth would suffer a higher frequency of catastrophic impacts.
4) A smaller frequency of catastrophic impacts increases the habitability of any planet.

   Hence,

5) Earth's habitability is increased by the presence of Jupiter in the Solar System.

This syllogistic formulation is useful for illuminating the issue of using the same labels for objects under at least controversial assumption that the terms of their reference stay the same in all parts of the argument. Horner and Jones investigate (and seem to refute) premise (2); but even if we retain it, the argument is incorrect, since the meaning of 'Earth' in (1) and (3) is different. In fact, premise (3) seems to be self-contradictory if we specify Earth as the planet we live on today and know reasonably well. From the purely physical point of view, there is no 'Earth without Jupiter', since by definition Earth is a planet formed and evolved through a complex historical process in which Jupiter played an important role. In the simplest possible form:

No Jupiter = no history of the Solar System as we know it = no Earth.

The two possible historical trajectories – the history of the Solar System with Jupiter and without it – are incommensurable. The contrary assumption – that *Earth* must be one and the same in (1) and (3) – is analogous to the "anachronistic assumptions" warned about in the historical context by Ferguson.[3] I shall discuss below whether modern metaphysics could come to the aid of Ward and Brownlee.

(A desperate proponent of REH could still claim that Earth is unimportant as a specific entity, and that we should anyway use it as a placeholder for something like 'a rocky planet in a habitable zone of its parent star'. However, this is self-defeating, since we lose the option of using observation selection to account for a-priori minuscule probability of finding ourselves on an 'Earth' once we find out that 'Earths' are incredibly rare. In other words, we need to account for the alleged increased habitability due to the presence of 'Jupiter' of this particular planet, i.e., our Earth, and not just any planet. I shall return to that point below.)

Unphysical *ceteris paribus* is usually a consequence of neglecting the causal context or oversimplifying the complexity of an observed situation. In supposing how the state of affairs could be different, Rare-Earth theorists assume simple, linear change, not taking into account the self-organizing nature of the relevant physical system, where a very small change at time t could cause dramatic divergence at some later time, $t + \Delta t$. Asking about the fate of Earth in the absence of Jupiter is self-contradictory from the point of view of conventional, mundane physics; Earth is a unique part of the complex system that includes Jupiter as a major component, so there are no guarantees that Earth would have existed at all if Jupiter were not present.[4]

Even if a similar planet existed (perhaps as a Lewisian "counterpart", see below), we would have to account for many other differences between that counterfactual situation and the actual one, so the question as to

3 I neglect here that part of the problem with REH is also that #4 (and analogous premises in other requirements) is uncertain – as explicated by Ward and Brownlee themselves. In other places (e.g., pp. 186-189) they claim that some impacts are in fact necessary for diversification of the biosphere. Obviously, the risk and habitability functions are more complex than intuitively assumed.

4 What if the alteration is located in sufficiently distant past? An amusing fictional example is provided by *A World of Difference* (Turtledove 2005). Notice that, for purely dramatic reason, while the main alteration occurs at the Solar System level, there is at least another, seemingly unrelated, political alteration considered in this novel.

what degree it is justified to call such a body 'Earth' would be very pertinent (see the section on transworld identity below). There is, of course, no way of magically removing Jupiter at some point in the history of the Solar System, but the conclusion of unphysical *ceteris paribus* would have stood even if there were such a supernatural occurrence. After all, the mechanism for enhancing Earth's habitability suggested by Ward and Brownlee is a protracted affair; even a magic-wand vanishing of Jupiter would not really influence our planet until a geologically relevant period of time has, say several million years, elapsed. Since most processes in the Solar System (and, indeed, anywhere in nature) are such that small uncertainties or perturbations in the initial conditions lead to exponentially divergent subsequent histories, the amount of changes caused by Jupiter vanishing after some relevant time would be huge; it is by no means clear whether the habitability of Earth would increase or decrease.

The same reasoning applies for some other REH requirements, notably the one dealing with the "rare Moon". As pointed out by Christopher Chyba and Kevin Hand in their fine review, the claim of chaotic obliquity fluctuations being supressed by the size of the Moon, leading to the habitable Earth of long rotational period we inhabit today, is something completely different from the counterfactual claim that the Earth would not be rotationally stable if the Moon never formed (Chyba and Hand 2005, pp. 47-48). Although it is difficult to estimate Earth's primordial rotation in the absence of a hypothetical Moon-forming impact, it is quite possible that a very small rotational period of the early Moon-less Earth would have ensured stable obliquity in any case. Again, the issue as to what extent would such a body, having a drastically different history (no large impact, no Moon, no subsequent tidal slowing down of fast primordial rotation), deserve to be called 'our' Earth is open to debate. Ditto for the consequent increase or decrease of habitability.

These are examples of a wider intellectual malady of our age: the tendency for over-atomizing complex things, searching for shortcuts through the complex causal webs of history, and imagining individual pieces as entirely changeable while magically preserving *ceteris paribus*. In a biological context, the same tendency has led to excesses of zealous ultra-adaptationism, and especially their misapplication in pop sociobiology and evolutionary psychology, which have been criticized by many (e.g., Kitcher 1985). It is the idea that advances in science are similar to putting together a children puzzle – starting from one corner

and advancing until just a few of the pieces are missing from the overall picture. However, this cute metaphor is a rather misleading account of the actual scientific process, for the shape of the puzzle picture, as well as the form of the remaining 'hole', changes continuously and as a function of the piece we are currently holding in our hand.

Transworld identity

How do we know what is Earth? We know for sure that it is a planetary body, but what other characteristics need to be specified to call such a body "Earth"? Could "Earth" be located in a different Solar System – specifically, in a Solar System with a different history – and still retain the label? This is an old and venerated issue in contemporary philosophy. As Mackie and Jago write in the relevant entry of the *Stanford Encyclopedia of Philosophy*:[5]

> Suppose that, in accordance with the possible-worlds framework for characterizing modal statements (statements about what is possible or necessary, about what might or could have been the case, what could not have been otherwise, and so on), we treat the general statement that there might have been purple cows as equivalent to the statement that there is some possible world in which there are purple cows, and the general statement that there could not have been round squares (i.e., that it is necessary that there are none) as equivalent to the statement that there is no possible world in which there are round squares.

Of course, this has been discussed most with respect to the identity of persons, but seemingly easier case of the identity of natural objects like planets is still highly controversial. While we may all safely agree that there are no round squares, what about non-habitable Earths? Is there a possible world in which Earth is not habitable?[6] We are accustomed to thinking about Earth as habitable and will even go so far to use Earth as the prototype of any and all habitable planets (as Ward and Brownlee consistently do in building the REH framework). But is that truly a *de re* statement? (The *de dicto* alternative would be that we are using terms "habitable" and "Earth-like" interchangeably, as synonyms.) And if it is a *de re* statement about particular individual – "our Earth" or "the Earth" – how are we to extend this to discuss *de re* modal versions of statements, like "Earth could not be habitable without Jupiter"?

5 Mackie and Jago (2013).

6 And here, of course, I do not consider Hadean Earth in the first 700 million years or so after its formation; instead, what we need to consider is the Earth at the same spatiotemporal and other relevant positions.

In the remainder of this paper I will use statements about Earth's habitability as *de re* statements, since if the research in astrobiology since 1995. shows anything, it is that under specific circumstances quite different astronomical objects, like Mars or Europa, could be habitable; this empirical content precludes the oversimplified *de dicto* reading. After we accept this, it is hard to overemphasize how difficult it is to discuss the modal *de re* statement without invoking or implying "magic wand" (i.e. unphysical) alterations of context. In the possible world semantics, the REH claim is equivalent to the claim that there is a possible world in which Earth exists without Jupiter (and is less habitable). From this it seems to follow that one and the same individual planet — *the* Earth — exists in some merely possible world, as well as in the actual world, so that there is an identity between Earth and some individual planet in another possible world. So, even if the original REH claim has *prima facie* nothing to do with possible worlds, instead talking about presumably actual Earth-like planets in our Galaxy, the argument involves a commitment to the metaphysical view that some individuals exist in more than one possible world, and thus to what is known as "identity across possible worlds", or *transworld identity* (Mackie 2006).

In itself, this is a remarkable conclusion, since it unexpectedly demonstrates the relevance of philosophy – and no less a discipline than *metaphysics*! – for a young scientific field such as astrobiology. In the course of the last century, the question of relevance of philosophy in the epoch of great strides of empirical science has often been posed, with assessments ranging from wildly negative and derogatory dismissals of philosophy, to reasonable calls for "peaceful coexistence" of science and philosophy. While the specific implications of the logical positivists' anti-metaphysical dogma have been downplayed in most post-positivist accounts, (at least since Quine), the dogma itself has hardly been criticized or undermined to this day. Here, however, we have a very specific example of a concrete scientific hypothesis (REH), whose claims cannot be properly evaluated without recourse to explicitly metaphysical assumptions such as those about transworld identity. Whether this is a freak example and exception confirming the rule or it points out in a new direction and offers unexpected prospects of fruitful collaboration of science and philosophy remains to be seen; some developments in theoretical cosmology, for instance, could be construed as a support to the latter view, but it remains to be seen.

What else remains?

If we reject the talk about transworld identity at all (which was, perhaps, Quine's point in asking whether the counterfactual bald man and the counterfactual fat man are one or two persons; cf. Quine 1953, 1976), than the answer to the problem of "rare Jupiter" becomes simple: there is no such thing as "Earth without Jupiter". Consequently, we cannot really hope to establish that other planetary systems are less habitable than the Solar System due to a lower measure of impact shielding. The amount of impact shielding in each particular case is a complicated function of the initial conditions which cannot be known in advance – and indeed cannot be calculated *at present* even for the perfectly known initial conditions (which we can never possess). In such a situation, the best course of action seems still to apply the Copernican principle – that is, contrary to Ward and Brownlee, to conclude that as to the amount of impact shielding corresponding to other planets in a planetary system, Earth is likely to be close to typical.

This seems counterintuitive, but we have already seen that intuition is prone to mislead us into taking REH claims more seriously than they can be justified. But, what about accepting other accounts of transworld identity, admitting that in at least some cases, we can justifiably speak of a "different" Earth in context of a different initial conditions and different planetary environment? Will those metaphysicists, more liberal in regard to questions of identity and modality, offer some quite unexpected help to Ward and Brownlee?

It seems that, under the present understanding (or else) of habitability, such help is not forthcoming. Take Lewisian modal realism, for example (Lewis 1971, 1986). It has been proposed to make account of otherwise plausible counterfactuals, such as "if I had not forgotten to water my plants, they would have not died." On the general form of modal realism, this implies that there is a world B (or a bundle of words B*), different from our world A, in which I have indeed watered the plants regularly in the past, and *as a consequence* they are alive at present. The tacit assumption is that we have a solid grasp of the causal relationship between watering plants and their survival. In spite of the fact that some preemption might occur – that, for instance, a catastrophic earthquake destroying my flat and the plants in it might have occurred yesterday in some of the other worlds, rendering the issue of plant survival moot – we safely assume that the general causal relationship between watering

plants and their survival is valid across possible worlds. This safe assumption is what makes the counterfactuals in question plausible. And in this example, it is a consequence of our *empirical knowledge of botanics* (which is essentially scientific knowledge, although in this particular example it goes far into the pre-scientific past so we usually do not think about as being particularly scientific).

However, the rare Earth/Jupiter counterfactual is not of such plausible kind. There is, to the best of our empirical knowledge of planetary and geosciences, no such safe causal relationship. We are, after all, dealing with probabilistic statements about chances for deflections and/or terrestrial impacts, spread over billions of years of strongly coupled nonlinear dynamical evolution. On the contrary, in recent decades we have more and more indications (not solid arguments, but useful indications nonetheless) that the Solar System is quasi-chaotic and self-organized physical system in which no simple causal relationship of the above variety holds. The REH counterfactual analog to the watering plant case would be something like the statement: "If it were raining in the jungle on a Thursday afternoon three months ago, this particular orchid would not have died." There is no safe causality here and no intuitive plausibility: the signal is drowned in the noise.

So we can see why modal realism and Lewisian counterpart theory cannot help much here. The standard interpretation of the counterpart relation is that an object X in world W_1 has as its counterpart in world W_2 whichever object is sufficiently like X *and* most like X. This implies that there must be a threshold of "sufficient likeness" for X in any other world; otherwise, X has no counterparts. In the specific case, we cannot judge, since we do not understand the relationship of habitability to other properties of Earth as a planet. We simply have no basis yet to establish the existence or else of Earth's counterpart in the Jupiterless world. Further research into the complex mess of causes and effects comprising the history of Solar System – and, presumably, *any* planetary system in any physically possible world – is necessary before we could pronounce on that issue. So the categorical REH claim fails.

Of course, it is still possible that other accounts of transworld identity might be more useful from the REH perspective. Suppose that we insist on calling "an Earth" any planetary body possessing a vague enough set of characteristics (implying some account of vague existence which is highly problematic in its own right, but which I cannot discuss here;

e.g., Smith 2005). Then, while it might formally be the case that the REH conclusion holds for a set of Earths, it would not really help in what is the intention of REH theorists: namely to show that the habitability of **our** Earth is highly improbable on other planetary bodies within our own world. Namely, the criteria for habitability in that case need to be at least as vague – which would more than compensate for the *a priori* improbability. In any case, it seems that more of both empirical work on disentagling various requirement for habitability *and* philosophical work is necessary before we could pronounce on this issue with more confidence. Simplicity of the original REH argument simply fails.

Discussion and prospects

We should not forget what counterfactual reasoning attempts to do when it is deployed as an explanatory strategy: the example of REH is quite typical and instructive in this respect. The main idea is to show that some particular aspect X of the observed reality has much smaller measure in the overall relevant parameter space than it is usually (or hitherto or *a priori*) assumed. After we perform counterfactual analysis of the relevant causal relations, we see that the Bayesian probability shift occurs and X is much less probable to occur and, indirectly, to be observed elsewhere. While historical counterfactuals have often not been so focused, it is clear that the quest for "crucial events" or "watersheds of history" has essentially the same structure: what distinguishes the observed historical trajectory from multiple might-have-beens are key acts of individuals or groups, which in themselves need not be of low probability, but their conjuction leading to *our* history is improbable. (This is analogous to the "rare Earth equation" of Ward and Brownlee.) In contrast to deterministes, proponents of Fergusonian virtual history hope to identify points where chance and subjectivity played important roles in the given historical trajectory by analyzing different logically consistent outcomes. In both cases, we could at least in principle and *sub specie aeternitatis* discuss probability of some particular feature of the observed world.[7] In the astrobiological context, if proponents of REH can constrain probability in some way, we could then pass on to the really focal question: applied to a set of similar locales in *our* world, how (im)probable are such features then? This is the question which could, at least in principle, be answered in an empirical manner.

7 Similar case is with evolutionary biology or paleontology: *what is the probability of evolving character X once or more times? Is the world without eukaryotes probable or not?* (cf. Gould 1989) – those are entirely legitimate questions to ask.

A great advance in the nascent philosophy of astrobiology has been introduced by the work of Iris Fry, who in several papers and an excellent book, *The Emergence of Life on Earth*, elaborated a key principle for the scientific study of biogenesis (Fry 1995, 2000). This principle she calls the continuity thesis:

> This common element, which I will coin "the continuity thesis", is the assumption that there is no unbridgeable gap between inorganic matter and living systems, and that under suitable physical conditions the emergence of life is highly probable. It is the adoption of the "continuity thesis", concerning the philosophical dimension of life-matter relationship, which has turned the origin of life into a legitimate scientific question, and which constitutes a necessary condition for any scientific research in this domain.

In essence, the continuity thesis encompasses almost all modern approaches to the origin of life (including those, *mutatis mutandis*, postulating the origin of life elsewhere and its transport to Earth via some form of panspermia). Its substance is evocatively summed up in the title of Chapter 13 of her book: "Neither by chance nor by design". Fry persuasively argues that the continuity thesis is the only meaningful way to proceed if one wishes to remain in the scientific domain; both the "lucky accident" naturalistic account of abiogenesis and supernaturalistic design lead us into the same blind alley.

The relationship to REH counterfactuals then emerges in a natural way. If we accept the continuity thesis, there is no reason to consider other habitable places in the universe as anything other than an extension of the habitable space on Earth, in terms of all particular properties of abiogenesis. In other words, we might establish *de re* criteria for habitability starting from "first principles" of physical, chemical, geological knowledge about a set of planets in the Galaxy. This would in turn enable locating Earth more precisely within the parameter space of habitability, thus obviating the need for counterfactual assumptions of untestable nature. Of course, such a research program will undoubtedly have many philosophical difficulties of its own; however, in light of its explanatory potential and great strides made in the last couple of decades on related issues, some optimism is warranted.

Acknowledgements. The author thanks Anders Sandberg, Jelena Dimitrijević, Branislav Vukotić, Richard Cathcart, Irena Diklić, Clement Vidal, George Dvorsky, J. Gaverick Matheny, Branislav K. Nikolić, Zoran Knežević, Nikola Božić, Mark A. Walker, John Smart, Dušan

Indjić, Srdjan Samurović, Seth Baum, Nick Bostrom, Martin Beech, and the late Robert J. Bradbury for useful discussions, kind support, and some technical help. This work has been supported by the Ministry of Education and Science of the Republic of Serbia through the projects ON176021 and ON 179048.

Primljeno: 27. novembra 2013.
Prihvaćeno: 17. decembra 2013.

References

Blumberg, B. S. 2011, "Astrobiology, space and the future age of discovery," *Philosophical Transactions of the Royal Society A* **369**, 508-515.

Chaisson, E. J., 2001, *Cosmic Evolution: The Rise of Complexity in Nature* (Harvard University Press, Cambridge).

Chaisson, E. J. 2003, "A unifying concept for astrobiology," *International Journal of Astrobiology* **2**, 91-101.

Chyba, C. F. and Hand, K. 2005, "Astrobiology: The Study of the Living Universe," *Annu. Rev. Astron. Astrophys.* **43**, 31-74.

Ćirković, M. M. 2012, *The Astrobiological Landscape: Philosophical Foundations of the Study of Cosmic Life* (Cambridge University Press, Cambridge).

Comins, N. F. 2010, *What If the Earth Had Two Moons?* (St. Martin's Press, New York).

Cowley, R. (ed.) 1999, *What If? The World's Foremost Military Historians Imagine What Might Have Been* (Bantam, New York).

Darling, D. 2001, *Life Everywhere* (Basic Books, New York).

Des Marais, D. J. and Walter, M. R. 1999, "Astrobiology: Exploring the Origins, Evolution, and Distribution of Life in the Universe," *Annu. Rev. Ecol. Syst.* **30**, 397-420.

Ferguson, N. 1999, *Virtual History: Alternatives and Counterfactuals* (Basic Books, New York).

Fry, I. 1995, "Are the different hypotheses on the emergence of life as different as they seem?" *Biology and Philosophy* **10**, 389-417.

Fry, I. 2000, *The Emergence of Life on Earth: A Historical and Scientific Overview* (Rutgers University Press, New Brunswick).

Gould, S. J. 1989, *Wonderful Life: The Burgess Shale and the Nature of History* (W. W. Norton, New York).

Grinspoon, D. 2003, *Lonely Planets: The Natural Philosophy of Alien Life* (HarperCollins, New York).

Haldane, J. B. S. 1927, *Possible Worlds and Other Essays* (Harper and Brothers, London).

Havel, I. M. 1999, "Living in Conceivable Worlds," *Foundations of Science* **3**, 375-394.

Hawthorn, G. 1991, *Plausible Worlds: Possibility and Understanding in History and the Social Sciences* (Cambridge University Press, Cambridge).

Heller, M. 2005, "Anti-essentialism and Counterpart Theory," *The Monist* **88**, 600-618.

Horner, J. and Jones, B. W. 2008, "Jupiter – Friend or foe? I: The asteroids," *International Journal of Astrobiology* **7**, 251-261.

Horner, J. and Jones, B. W. 2009, "Jupiter friend or foe? II: The Centaurs," *International Journal of Astrobiology* **8**, 75-80.
Impey, C. (ed.) 2010, *Talking About Life: Conversations on Astrobiology* (Cambridge University Press, Cambridge).
Kitcher, P. 1985, *Vaulting Ambition: Sociobiology and the Quest for Human Nature* (MIT Press, Cambridge).
Lewis, D. 1971, "Counterparts of Persons and Their Bodies," *The Journal of Philosophy* **68**, 203-211.
Lewis, D. 1986, *On the Plurality of Worlds* (Blackwell, Oxford).
Mackie, P. 2006, *How Things Might Have Been: Individuals, Kinds, and Essential Properties* (Oxford University Press, Oxford).
Mackie, P. and Jago, M. 2013, "Transworld Identity", *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2013/entries/identity-transworld/>.
Néron de Surgy, O. and Laskar, J. 1997, "On the long term evolution of the spin of the Earth," *Astronomy and Astrophysics* **318**, 975-989.
Quine, W. V. O. 1953, "On What There Is," in *From a Logical Point of View*, (Harvard University Press, Cambridge, Mass.).
Quine, W. V. 1976, "Worlds Away," *The Journal of Philosophy* **73**, 859-863.
Smith, N. J. J. 2005, "A Plea for Things That Are Not Quite All There: Or, Is There a Problem about Vague Composition and Vague Existence?" *The Journal of Philosophy* **102**, 381-421.
Toynbee, A. J. 1969, *Some Problems of Greek History* (Oxford University Press, Oxford).
Turtledove, H. 2005, *A World of Difference* (Del Rey Books, New York).
Ward, P. D. and Brownlee, D. 2000, *Rare Earth: Why Complex Life Is Uncommon in the Universe* (Springer, New York).

Milan M. Ćirković

Protivčinjenični iskazi i nefizički ceteris paribus: jedna eksplanatorna greška

Rezime

U ovom radu razmatram tip protivčinjeničnih argumenata često korišćenih u istorijskim naukama na mnogo diskutovanom skorašnjem primeru „hipoteze retke Zemlje" u planetarnim naukama i astrobiologiji. Argument se zasniva na navodnoj „retkosti" nekih od ključnih komponenti planetske nastanjivosti koje, u slučaju Zemlje, stvara njen kontingentni istorijski razvoj. Na primer, tvrdnja da kontingentna istorijska činjenica postojanja planete Jupiter omogućuje Zemlji zaštitu od katastrofalnih sudara i tako povećava Zemljinu nastanjivost, često dovodi do zaključka da takvo stanje stvari mora biti veoma retko u našoj Galaksiji. Ovde pokazujem da je to rasuđivanje duboko pogrešno, zbog nekoliko blisko vezanih razloga. Uz to, ispostavlja se da je filozofski problem transsvetskog identiteta veoma relevantan za ovu vrstu istorijskog rasuđivanja u empirijskim naukama. Ovaj primer pokazuje brojne eksplanatorne problema sa kojima se suočavamo u pokušaju korišćenja istorijskih protivčinjeničnih iskaza u proučavanju kompleksnih, nelinearnih dinamičkih sistema – ali i daje podstreka relevantnosti filozofije za evaluaciju eksplanatornih tvrdnji u posebnim naukama.

Ključne reči: protivčinjenični iskazi, astrobiologija, filozofija nauke, filozofija istorije, transsvetski identitet