



What good are abstract and what-if models? Lessons from the Gaïa hypothesis

Sébastien Dutreuil

► To cite this version:

Sébastien Dutreuil. What good are abstract and what-if models? Lessons from the Gaïa hypothesis. History and Philosophy of the Life Sciences, 2014, 36 (1), <http://link.springer.com/article/10.1007/s40656-014-0003-4>. 10.1007/s40656-014-0003-4. hal-01109171

HAL Id: hal-01109171

<https://hal.science/hal-01109171>

Submitted on 25 Jan 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

What good are abstract and what-if models? Lessons from the Gaïa hypothesis.

A final version of this work has been published. Please, cite the published version, available here:

<http://link.springer.com/article/10.1007/s40656-014-0003-4>

Dutreuil, Sébastien, 2014, “What good are abstract and what-if models? Lessons from the Gaia hypothesis?”, *History and Philosophy of the Life Sciences*, Volume 36, Issue 1, pp.16-41, DOI 10.1007/s40656-014-0003-4

Sébastien Dutreuil

Université Paris 1 Panthéon-Sorbonne

UMR 8590 IHPST – Institut d’Histoire et de Philosophie des Sciences et des Techniques

13, rue du Four

75006 Paris

+33671755851

seb.dutreuil@gmail.com

Abstract

This article on the epistemology of computational models stems from an analysis of the Gaïa Hypothesis (GH). It begins with James Kirchner's criticisms of the central computational model of GH: Daisyworld. Among other things, the model has been criticized for being too abstract, describing fictional entities (fictive daisies on an imaginary planet) and trying to answer counterfactual (what-if) questions (how would a planet look like if life had no influence on it?). For these reasons the model has been considered not testable and therefore not legitimate in science, and in any case not very interesting since it explores non actual issues. This criticism implicitly assumes that science should only be involved in the making of models that are "actual" (by opposition to what-if) and "specific" (by opposition to abstract).

I challenge both of these criticisms in this article. First by showing that although the testability – understood as the comparison of model output with empirical data – is an important procedure for explanatory models, there are plenty of models that are not testable. The fact that these are not testable (in this restricted sense) has nothing to do with their being "abstract" or "what-if" but with their being predictive models. Secondly, I argue that "abstract" and "what-if" models aim at (respectable) epistemic purposes distinct from those pursued by "actual and specific models". Abstract models are used to propose how-possibly explanation or to pursue theorizing. What-if models are used to attribute causal or explanatory power to a variable of interest. The fact that they aim at different epistemic goals entails that it may not be accurate to consider the choice between different kinds of model as a "strategy".

KEYWORDS : Gaïa hypothesis ; Computational Models ; Daisyworld ; Artificial Life; Explanation; Model Validation

Introduction

The Gaia hypothesis (GH) was originally proposed by Lovelock and Margulis (1974) in a foundational article entitled “Atmospheric homeostasis by and for the biosphere: the Gaïa hypothesis” in which the authors suggested that the influence of life on the geological environment may contribute to a global homeostasis, which lead them to compare the ensemble comprising all the living beings and the environment with which they interact to an organism. The fact that the Earth does not reproduce and can not undergo natural selection was pointed out as problematic for the emergence of homeostasis at this level. It was indeed argued that it is because genuine organisms can undergo natural selection that they exhibit homeostasis and then claimed first that Earth cannot undergo natural selection because it does not reproduce, second that appealing to natural selection at lower levels would still carry the theoretical problems of group selection (Dawkins, 1982, 235–237). Given these theoretical difficulties, GH teleological claims are not justified. Ever since, philosophers of biology turned their back to what appeared to be nothing more than a misleading metaphor.

The fact that GH has been considered to be very controversial by the vast majority of biologists and philosophers of science makes it very interesting from a philosophical point of view: the intensity of the debate has led to sharp discussion and critiques of several epistemological issues. James Kirchner has undoubtedly made the most informed critiques of GH. He greatly contributed to a clarification of its *explanandum* and confronted GH with a number of interesting epistemological problems: the role of metaphors in science, the demarcation between science and pseudo-science, the nature and roles of models (Kirchner 1989).

A substantial scientific literature, mostly led by geoscientists (Lovelock, Watson, Lenton, Volk, Kleidon, Kirchner, etc.) but not only (Margulis, Wilkinson), starting in the 1980's and

expanding in the late 1990's, was dedicated to the development of GH. After the early critiques, several directions of research have been envisaged, from empirical to more theoretically orientated ones, that took either the form of verbal arguments or of computational models. It is worth noting that the very first answer to critiques did not take the form of verbal arguments addressing the status of natural selection, or of the import of elements from general philosophy of science (those came later in the debate), but was the publication, by Watson and Lovelock (1983), of a simple computational model: Daisyworld.

This model was criticized by Kirchner (1989, 2002, 2003) on scientific and epistemological grounds. At the scientific level, he claimed, the hypotheses of the original model were too strong. I will not discuss in details all the pertinent scientific critiques raised by Kirchner and partly answered afterwards.¹ The second kind of critique that Kirchner raised is an epistemological one. It dismisses the *model* Daisyworld – or rather, as I will show, the *questions* Daisyworld is meant to answer – as non-scientific at worst, and not interesting at best. Contrary to the scientific critiques, epistemological ones have not been taken so seriously by the authors of GH.

Rather than beginning with general philosophical considerations over the status of models in science and then evaluating how they could be used or applied in the context of GH, this article will proceed the other way around. It starts with a brief overview of GH to understand where the debate on models originated from. It then makes explicit the critiques addressed to Daisyworld by elaborating a typology of different kinds of models used in the context of GH. The typology contrasts the supposedly wrong kinds of model (“abstract” and “what-if”) with

¹ For reviews, see Wood et al. (2008), McDonald-Gibson et al. (2008), Lenton and Williams (2009), Dutreuil (2013).

allegedly better kinds of model (“actual and specific”). What good are abstract and what-if models? The question is twofold: (i) are “abstract” and “what-if” models legitimate in science? And, (ii), are they useful? *Contra* Kirchner, I will argue that the answer to both questions is a positive one, i.e. that the problems faced by GH are not epistemological (and therefore do not concern the legitimacy of GH’s kinds of question or model); rather, the problems encountered by GH models, if any, are conceptual ones (they concern the meaning of some terms) or scientific ones (the hypotheses made by the models).

The typology thus elaborated will prove to be fairly general and the different types of models studied will be shown to be pervasive across the sciences. The discussion over the epistemological status and interest of abstract and more particularly of “what-if” models aims at contributing to recent philosophical discussions over the status of models in science. So the overall ambition of this paper is twofold. It first aims at clarifying the debate over the status of some of the models designed by GH authors: as such, it tackles an issue that is internal to GH. But it also strives to extend its scope beyond GH to touch upon philosophical considerations over the status of models in science more generally.

1. Daisyworld: The model and its critiques

GH, often misleadingly summarized as a metaphor comparing the Earth to an organism, is not commonly thought to be worth consideration by the scientific community, and even less so by philosophers of science.² However, carefully (or bluntly) dismissing the question “Is Earth

² Sarkar and Plutynski’s judgement of GH as being a “philosophically intriguing idea *at the fringe of science*” (2010, p. xxiii, my emphasis) belongs to the more sympathetic

‘really’ alive?’ as not interesting is not enough to reject GH. This question as such may not, indeed, be interesting; more generally it may also be that the broader question “is X really alive” is not interesting and that one should focus on more specific properties than “being alive”, as Sober (1991) has argued. But it is also *not* the question that GH raises. Neither does GH consider that invoking the fact that Earth is alive provides any meaningful explanation of any biological or geological phenomenon.

Instead, as I have argued elsewhere, drawing in part on Kirchner (1989), GH focuses on three different questions (Dutreuil 2012):

- (i) Does life *influence* geological phenomena (climate, the composition of the atmosphere and oceans)?
- (ii) Can the influence of life on its environment affect Earth’s *habitability*?
- (iii) Can life’s activity *regulate* environmental variables?

Question (i) is unproblematic *per se*. If there is a contention, it belongs to the history of science and regards the role that GH had in drawing attention to the influence that life may have on its environment. The crux of the matter lies in question (ii) and (iii) and in their distinction; both can be understood as instances of a broader question: “Can life benefit from its influence on the environment?” which is central to GH and to a later movement in ecology

positions within the community. According to Ruse (2013), part of the reasons for the rejection of GH by evolutionary biologists, besides epistemological issues, are to be found in the own insecurity and tension that existed within evolutionary biology at the time GH was proposed (chap. 8).

and evolutionary biology, the niche construction literature.³ Maintenance of habitability is obviously also a question of regulation; so question (iii), in that it deals with regulation *per se* – and not with regulation as a mean to achieve habitability – is tied to stronger claims of GH attributing to life's influence the capacity to make the environment *optimal* for life. In conclusion, GH's explanandum cannot be reduced to the question of whether one can genuinely compare Earth with an organism.

Surely, everyone agrees that it was pertinent to point out early, as Doolittle (1981) and Dawkins (1982) did, that a mechanism responsible for the maintenance of homeostasis at a planetary level was required, given the fact that natural selection does not operate at this level. But then a reasonable philosophical and scientific attitude consists in paying attention to the propositions that are made or to the explanations that are offered, and not in dismissing the whole GH literature as unscientific for asking a question it does actually not ask.

Since the early 1980's, then, the central problem of the GH has been to find a mechanism that could maintain homeostasis at the scale of the Earth. Watson and Lovelock (1983) made one move in this direction. And again, they did not do so through a conceptual analysis of the notion of reproduction or theoretical refinements on the conditions under which natural selection occurs: they built a computational model, Daisyworld.

The original model stages what the authors describe as an "imaginary planet" where the evolution of temperature depends on only two parameters: average planetary albedo and an external forcing, a gradual increase of solar luminosity. The planet is populated by two species of daisy: black ones have a lower albedo (0.25) than the bare ground (0.5) and white

³ On GH and niche construction, see Free and Barton (2007), Pocheville (2010).

ones have a higher albedo (0.75). Therefore the relative proportion of white and black daisies (covering the ground) influences the planet's albedo, and hence temperature. In return, the temperature constrains the growth of daisies. So life affects an environmental variable that is affected by an external perturbation and that constrains life's growth in turn. The model is fairly simple, consisting of a few differential equations, and does not explicitly describe space (see fig. 1.a). The central conclusion drawn by Watson and Lovelock is that "regardless of the details of the interaction, the effect of daisies is to stabilise the temperature"; and this occurs without any form of selection acting at the level of the system itself.

There are three important features of the Daisyworld model. The first interesting aspect of Daisyworld is what I will temporarily call its "*fictional*" dimension. Recall that the title of Watson and Lovelock (1983) was "Biological homeostasis of the global environment: the *parable* of Daisyworld" (my emphasis) and that the model was taken to describe an *imaginary* planet. Many subsequent Daisyworld papers used words pertaining to the semantic field of fiction when talking about their model: "parable", "imaginary planet", "toy model", "toy world", "a caricature of the Earth system", "fictitious Daisyworld", "metaphor".⁴

Although this vocabulary is common in the modelling literature, it was criticized by Kirchner in the following words:

Building and testing quantitatively appropriate models of *real-world* atmosphere-biosphere interactions, although harder work than building *hypothetical* Daisyworlds, is likely to yield more relevant insights for Earth system science. Recent modelling efforts by Lenton and colleagues (Lenton, 2001; Lenton and Von

⁴ For a detailed review, precise quotations, references and pages, see Dutreuil (2013).

Bloh, 2001; Lenton and Watson, 2000) are a step in the right direction. (Kirchner, 2003, 40–41)

There is no such thing in the world as this simple planet populated by two population of giant daisies. It seems, then, that Daisyworld describes and strives to understand the behaviour of *fictional* objects (as opposed to *actual*). Should science take time and energy to understand objects that do not exist and if so, why?

The second interesting aspect of Daisyworld is its *abstract*, as opposed to *specific*, nature. The abstract/specific distinction pertains to the level of description. Specific questions bear on particular processes while abstract ones focus on properties that are shared by different processes. Kirchner questions the relevance of models that would be too abstract:

Second, it is important to move beyond simply theorizing. [...] After several decades and several iterations of Gaia theory, there is an urgent need for that theory to be *tested against data*. Unfortunately, Gaïa theory makes *predictions* that are abstract and qualitative, making them difficult to test in the real world. (*Ibid.*, my emphasis)

The test of Gaïa theory and models against data has been one of the principal concern of Kirchner, as one can see by looking at the title of its principal contributions: “The Gaïa hypothesis: can it be tested” (1989), “The Gaia hypothesis: fact, theory, and wishful thinking” (2002), “The Gaïa hypothesis: conjectures and refutations” (2003). I acknowledge the important contributions of Kirchner when he tries to move the discussion toward an assessment of the empirical relevance of some claims of GH (Kirchner 2002, 395–398, Kirchner 2003, 26–29). However, his claim of the impossibility to test the conclusions of Daisyworld *qua* model (1989, 230), or, more recently *qua* “abstract model” (2003, 38),

deserves further discussion over the meaning of “abstract” and over the way models are constrained by the empirical world.⁵

The third and last characteristic of Daisyworld is the exploration of counterfactual worlds, worlds that contradict in relevant respects certain properties of the actual world; this exploration relies on what I will call “what-if” questions. As Kirchner (2003) noted:

Many of the proposed tests for Gaïa are based on assessing whether Earth’s surface environment would be different without life. Would an abiotic Earth be less resilient and resistant to perturbation (Lenton, 2002)? Would the environmental conditions prevailing on an abiotic Earth be less conducive to biological productivity (Kleidon, 2002)? These hypotheses have the advantage of being fairly precise, but the disadvantage of being untestable in practice. They also have the more serious disadvantage of being far removed from the most pressing issues in Earth system science – namely, how the climate system works and how it will evolve in the future. (*Ibid.*)

We have drawn three distinctions so far: fictional/actual, abstract/specific, what-if/actual. Note that the abstract/specific distinction is orthogonal to the two others. One can for instance explore “what-if” questions at a specific level of description (what would be the actual temperature of Earth if angiosperms had not arisen and influenced erosion processes?) or at an

⁵ Kirchner also worries about the *simplicity* of Daisyworld which may not reflect the complexity of the world; for a discussion of this point, see Dyke and Weaver (2013), Dutreuil (2013).

abstract one (how would “environment” differ if “life-in-general” was not to affect it?). We will see later in what regard “what-if” differs from “fictional”.

Kirchner brought to the fore very interesting epistemological questions, on which Lovelock himself has not been very clear, not to say contradictory (Kirchner 2003, 37) or purposefully provocative. However, I think that some of Kirchner claims are disputable and that, in any case, these questions deserve more space and more discussion than Kirchner’s account. Kirchner did not clearly draw the above distinctions. Nonetheless, it is clear that he is suspicious of the terms on the left (fictional, abstract, what-if). The first reason for this is that these models are (allegedly) about something else than the “real” world; in reading Kirchner, it seems that “real” means *specific* and *actual*. I do not want here to engage in metaphysical debates – do abstract phenomena exist? are counterfactual worlds less real than the actual one? – but will instead examine whether science should engage (and if so, why) in the study of abstract or counterfactual worlds. The second cause for worry for Kirchner is that abstract or what-if models are not “testable”, which is problematic for the Popperian stance ostensibly adopted by Kirchner. Interestingly, Kirchner opposes the questionable models with what he thinks to be a better practice: recent models from Lenton and colleagues, who propose specific models of the actual climate.

Two different questions are thus raised: (i) are fictional, abstract and what-if models epistemologically sound: are they scientific? are they useful? (ii) Are they less useful than other kinds of model? In the following, I will show that they are scientific and useful, and also briefly touch on the latter question.

3. Description of the models

Before engaging in a close description of Daisyworld, and other, related models, I want to briefly discuss what models are. One makes a model of a target system in order to gain knowledge, by explanation or prediction, of this particular target system. Making a model can be thought of as a particular strategy within science at large (Godfrey-Smith 2006, Weisberg 2007). Regardless of the conception of model that one adopts, two central elements will be preserved: models are *representations* and models rest on an *analogy* between the model system and the target system (the model system must resemble in relevant ways the target system). A famous example, discussed by Weisberg, is the Lotka-Volterra system of equations (the model system) representing prey/predator interactions between fishes in the Adriatic sea (the target system). Models can be implemented in different substrates: some models rely on concrete physical objects, others are mathematical, others computational. All models discussed here are computational. Although computational models are based on mathematics, they can be (and should be) distinguished from purely mathematical models (e.g. Lotka-Volterra equations) since they function differently in practice (Weisberg, 2013, 20) and since they raise particular epistemic and philosophical issues in that they are computational and not only in that they are models (Humphreys 2009, Grüne-Yanoff and Weirich 2010).

I pointed out above that the lexical field of fiction was common in the scientific literature on models. This was in fact the starting point for recent accounts which treated mathematical models as fictional entities (Godfrey-Smith 2009, Frigg 2010). This view is opposed to the one that takes mathematical models to be abstract mathematical entities (Weisberg 2013, ch. 3 and 4). The debates thus bears on what mathematical models *are* and not what they are *about*, or what the nature of their target system is.

Whether one conceives model systems as abstract or fictional may lead one to face different difficulties regarding *how* knowledge can be attained by them (Levy 2012). The debate just mentioned thus concerns the ontological status of the *model system*, an issue on which I will remain neutral. Instead, I will use “abstract”, “fictional” and “what-if” as terms qualifying the *target* of the model, the system the modeller wants to gain knowledge about. This, I think, does more justice to Kirchner’s discussion. Kirchner has not dismissed *all* kinds of model as works of fiction, he has only dismissed those that seem to be *about* abstract or counterfactual entities.

In the following, I will first describe in detail one model described as “a step in the right direction” by Kirchner (2003, 39) and then turn to examples of the abstract and what-if models that he deemed problematic.

3.1 “Good” models: specific and actual

In 1982, one year before the publication of Daisyworld, Lovelock and Whitfield published a paper entitled “Lifespan of the biosphere”. Their discussion was rooted in the so-called “faint young sun paradox”: climate on Earth seems to have remained stable in the last 4.5 Gyr even though solar luminosity has kept increasing. Lovelock and Whitfield suggested that life may have had an important role to play in fostering silicate erosion, which would have lowered the partial pressure of CO₂ (pCO₂) during the past billion years, balancing the increase of solar luminosity. However, since pCO₂ is already very low now, if no other balancing mechanism takes place, temperature will steadily increase in the next hundreds millions of years, as solar luminosity keeps increasing. The biosphere is thus threatened either through a long-term increase of temperature, or through CO₂-starvation, which raises the question of its future lifespan.

Lenton and von Bloh (2001), cited by Kirchner as a “good” model, is one of the five other models that followed on the topic (Caldeira and Kasting 1992; Franck et al. 2000, 2006; Li et al. 2009). All models aim at answering the question “How long will the Earth remain habitable?”. These five models thus need to do two things: (i) evaluate what the most relevant environmental variables constraining the survival of *actual* living beings are; and (ii) quantify the evolution of the variables that are important to sustain life. As to (i), the answer importantly depends on the living beings taken into account: *photosynthetic* life may disappear because of CO₂-starvation if pCO₂ falls below a certain threshold, *complex* life may disappear because of too high a temperature (greater than 50°C), and in any case *no known* form of life will remain when the temperature will be so high that water will have escaped Earth. As to (ii), the simulation calculates the evolution of pCO₂ and of mean temperature (two linked variables) based on a simple climate model and on other parameters: solar luminosity predicted from models of stellar evolution, weathering rate, etc.

3.2 Two kinds of abstract models

There are two kinds of abstract models, “abstract *simpliciter*” and “abstract and fictional”. “Abstract *simpliciter*” models are exemplified by the FLASK model, first published by Williams and Lenton (2008). They describe the model as follows:

[It] simulates an evolving population of microbes suspended in a flask of liquid, and hence the name. There is a prescribed supply flux of different nutrients into the flask and corresponding removal fluxes proportional to the concentration of each nutrient in the flask. There are also non-nutrient “abiotic” environmental variables. The flask is seeded with a clonal population of model microbes [...]. (Lenton and Williams 2009, 67)

This model seems to be describing things that one could observe in the actual world (flasks of microbes). However, it becomes clear when reading the article in detail that what really interests the authors (the scope of the model) is *the abstract properties* of the living beings represented: the authors ask us not to focus on the specific properties of the flask by giving arbitrary names to the nutrients (N1, N2, ...). Likewise, they abstract away from a specific strain of microbes, taking the “microbes” to be exemplars of what it is to be a living organism in general. The model is thus presented as giving us a potential mechanism for homeostasis to emerge at the scale of the Earth. The target system of this model, and of other variants of Daisyworlds that have been published, are the abstract properties of living beings and the consequences of these properties on the environment. This has been made explicit in statements such as: “organisms have an effect on their environment”, “organisms can only survive in a certain environment”, “there are constraints or limits on adaptation” (McDonald-Gibson et al. 2008, 654). Such statements illustrate what Weisberg calls “generalized models”:

A generalized model of sexual reproduction isn't supposed to be about kangaroo sex or fungi sex, but about sex itself.[...] But nothing in the world looks like “sex in general”. There is kangaroo sex, Tasmanian devil sex, and human sex, but not sex in general. Sex in general is an abstraction over these more specific kinds of sex. (Weisberg 2013, 116)

Whereas it mattered for the specific models described in the previous section that the particular molecule consumed by photosynthetic activity was CO_2 (because CO_2 was involved in other specific relations to other items of the model), for abstract models, the particular nature of the nutrient and waste will be irrelevant: only the fact that all living beings take up nutrient and excrete waste will matter. To be sure, all models may be thought to imply

a degree of abstraction; this is after all what makes science interesting: it abstracts away from local, individual, particular, specific phenomena. Moreover, the specific/abstract distinction comes in degrees, contrary to the actual/what-if, or to the actual/fictional distinctions: a class of properties may be more or less abstract.

In Weisberg's view of abstract or generalized model, in some cases the target system and the model system have the same level of abstraction: the model represents exactly the abstract features of interest of the target system, no less, no more. This would be the case for what I called the "abstract *simpliciter*" models, described above. What I mean by "abstract *simpliciter*" – models in which the target system and the model system have the same degree of abstraction – will be better understood by contrast with models that are *not* "abstract *simpliciter*", namely "abstract and fictional" model. In the latter case, for different reasons, the modeller does not build a model that is exactly as abstract as its target system: the model is more specific than the abstract properties of interest.

A very good example of such "abstract and fictional" models is provided by the original Daisyworld paper. As Watson and Lovelock (1983, 284) put it, "We are not trying to model the Earth but rather a fictional world which displays clearly a property which we believe is important for the Earth". Their model proceeds in two steps: (i) it abstracts away to important properties of living organisms and their planetary environment (organisms are affected by and affect in return their environment which is besides constrained by an external forcing), (ii) it imagines *fictive* with specific properties coherent with the abstract properties identified before. What distinguishes abstract *simpliciter* from actual and fictional model is that in the latter case some properties of the entities described verbally are actually implemented in the model and specified in the equations (here through the parametrisation of the albedo of the daisies, their preferred temperature, and so on). The intention of the authors is not to gain

knowledge about the speculative Daisyworld, it is to draw lessons from the analysis of Daisyworld, that can be applied to Earth. This can only be done on two conditions: first if, in the given context of inquiry, the relevant abstract properties of the model system are similar enough (given one's fidelity criteria) to the abstract properties of the target system (Earth and its living beings); second if the authors clearly define the intended scope of their model.⁶ So in the original Daisyworld, the model is much more specific (because of all the details about the daisies) than its target system (abstract properties common to all living beings). For these reasons, Daisyworld is a good example of what I called "abstract and fictional": its *relevant abstract properties* resemble that of our actual world while *irrelevant specific properties* (the parametrisation of daisies albedo and growth) differ from it.

One question still remains: should one adopt "abstract *simpliciter*" rather than "abstract and *fictional*" models in a particular context of inquiry? It is interesting to take a look at the historical path that led, in the case of the GH, from the latter to the former. About a hundred variations of Daisyworld have been published since 1983. Some of these variations progressively gave up on important elements of the Daisyworld style (weird daisies and specific details) and became more explicitly "abstract *simpliciter*". This was done in part by simplifying the initial model and by focusing more on the abstract properties of the entities (Harvey 2004, Williams and Noble 2005, Dyke and Harvey 2006, McDonald-Gibson et al. 2008). This was also done through the initiation of new kinds of models getting rid off the daisies and hinging on the *Artificial Life* (A-life) tradition, such as Downing and Zvirinsky (1999). The FLASK model hails from this A-life branch. It is interesting to point out that

⁶ For details about "fidelity criteria" and "intended scope", see Weisberg (2013), section 3.3.

despite fictional details being progressively abandoned by some models, others kept the “abstract and fictional” style, such as Boyle et al. (2011) or Weaver and Dyke (2012). We will discuss whether one should choose between “abstract and fictional” and “abstract *simpliciter*” at the end of the paper, as it is related to the epistemic purposes of the models.

3.3 What-if models

What-if models rely on the comparison of two different scenarios. Here I use scenario to denote the outcome of a model in a given configuration. What can be changed in a model to produce different scenarios is the values of the initial parameters and, to some extent, part of the model structure (the relation between some properties of the model). The main figure of the original Daisyworld paper shows the evolution of the temperature in two different scenarios (fig. 1a): one in which the model is run in normal or actual mode, and one in which the influence of life on the environment is artificially switched off (the “what-if” scenario: “what if life had no influence on the environment?”). In this case, a *relevant* parameter of the model (the influence of life) is artificially switched off in order to compare the outcome of the actual scenario with that of the what-if scenario. A quite similar approach can be found in experimental biology when one knocks off a gene of a model organism and compares the phenotypes of the mutants and wild type. So what-if scenarios differ significantly from “fictional” worlds since fictional models represent entities that differ from the actual world with respect to *irrelevant* specific properties but resemble it with respect to relevant abstract properties; by contrast the what-if scenarios differ from the actual world for *relevant* (supposedly the most relevant) properties.

I called “what-if model” the practice consisting in comparing the outcome of the same model run in two different scenarios (one actual, one “what-if”). Here we have thus two pertinent units of analysis corresponding to two different practices of modelling: the

scenario and the model. The scenario and the model differ regarding their epistemic purposes. Both can be interesting as such : in the case of the what-if scenario, one is genuinely interested in something else than the actual world whereas in the case of the what-if model, one is interested in the particular role played, *in the actual world*, by the principal variable of interest, namely the one that is switched off or changed in the “what-if” scenario. The what-if model constitutes an autonomous unit of description, distinct from the scenario, in that what carry the explanatory power is the comparison between two different scenarios.⁷ In all the GH cases discussed, what-if scenario are never explored as such but always in a context of a what-if model. So, clearly, the question that interests the authors is the role played by life in the *actual* world.

⁷ One may worry that it may not be warranted to talk about *one* model and to talk about the *same* model (run in two different scenarios) since not only the values of some parameters may be modified but also the structure of the model. First, the modification of the structure that takes place here can actually be thought of as a borderline case of the change of the value of some parameters (bringing one parameter to zero). Second, calling model-A and model-B the two different scenarios does not change my argument, namely the idea that the comparison between the two (models or scenarios) is a particular modelling practice, which, as it will be argued below, is epistemologically sound and has a particular epistemic purpose.

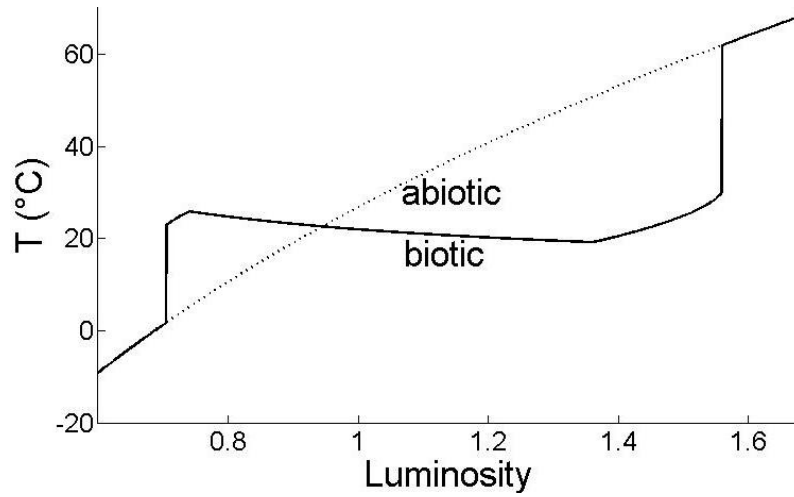


Fig. 1: a. Personal reproduction of Daisyworld central figure, first published in Watson and Lovelock, 1983, fig. 1, d, bottom. The figure has been reproduced based on the equations of Watson and Lovelock 1983 and on the procedure instructions given in its section 2 and 3. The legends “biotic” and “abiotic” have been added, based on Watson and Lovelock 1983 discussion in the text.

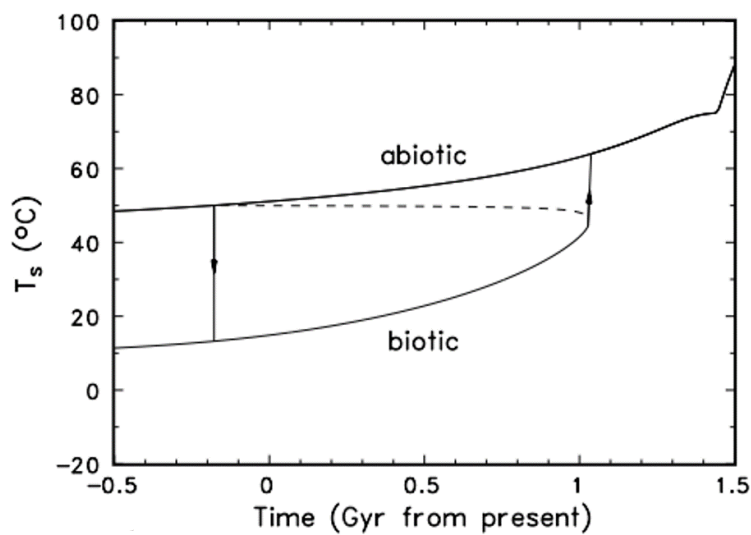


Fig 1. b. Reproduction of Lenton and von Bloh (2001) figure (1. a., in the original article). All curves show the evolution of temperature as solar luminosity (time) increases. Both simulations are run in two different scenarios: one in which the influence of life on the

environment is switched on (biotic), another one in which it is switched off (abiotic: the “what-if” scenario defined in this paper).

Some what-if models are also abstract models such as – FLASK and Daisyworld – , others are specific ones such as the model proposed by Lenton and von Bloh that I discussed above (2001; see fig.1b) : the distinction between what-if and actual is thus orthogonal to the distinction between abstract and specific. Besides, authors not involved in the development of GH have adopted similar strategies, wondering for instance how Earth would look like had Earth’s geomorphology not been influenced by life (Dietrich and Perron 2006); the making of what-if models is therefore not an oddity specific to GH modellers.

4 Discussion of the models

After having tried to clarify what I mean by abstract and what-if models I will first address the issue of their trustworthiness and legitimacy. I will then proceed to discuss the epistemic roles) of each model, and, based on this discussion, question the idea that the choice between these different kinds of model should be construed as a strategy.

4.1 Trustworthiness and legitimacy of the models

In order to address the issue of the legitimacy and trustworthiness of computational models it is useful to start with Kirchner’s remarks on their testability. Two senses of “testability” can be distinguished in his 2003 article. One seems to be restricted to the comparison of model *predictions* with empirical data: I shall refer to this use of the term as *testability_O* (for “output”). The second sense is broader and close to Popper’s criterion of demarcation between science and pseudo-science. Although Popper has remained popular among scientists (Pigliucci 2008), it is worth remembering that Popper’s demarcation criterion (falsifiability) has been seriously put into question for being too narrow in some cases and too wide in others

(e.g. Lakatos 1978); its pertinence has in particular been questioned when applied to the intelligent design debate (McCain and Weslake 2013, Sober 2008, 128–131). I will not linger on such general philosophical issues and will choose to understand Kirchner’s appeal in a simpler and uncontroversial way, namely that models should at some point be constrained by the “real” world, at least if they are to tell us something about the world.

This being granted, it should not be forgotten that important sources of reliability in a model have nothing to do with empirical data, or with the “external world”: they are internal to the model itself (Oreskes et al. 1994, Barberousse and Vorms 2013, Winsberg 2006). Verification of numerical solutions (through benchmarking procedures) is a good example of procedures that are important for the reliability of a model; checking the good behaviour of the numerical simulation after the integration of all the algorithms (routines) is another example. However, these purely computational and mathematical aspects, though important for the reliability of a model, are not our principal matter of concern: what we want to know is how the external world comes into play with a model, how it constrains it in relevant ways so that the model can tell us “something” about the world. The answer to this epistemological question crucially depends on what the “something” is that the model is supposed to tell us. In other words, the way the external world will constrain our model ultimately depends on the reasons for which the model was made.

So why does one make computational models? There are two answers to this question: to *explain* or to *predict*. A computational model can be construed as a tripartite object comprising a *model structure* (defining relations between properties) that is fed with *input values* (initial conditions of variables and parameters, that constitute different scenarios) in order to give us *output values* (the model predictions). Hence, there are two things in which one may be interested, two reasons why one may build the model in the first place and,

accordingly, there are two different ways the external world may come into play. In the case of *predictive* models one relies on the model structure to obtain predictions only. Meteorological models constitute a good example: one relies on the model structure (based on the laws of thermodynamics and of fluid mechanics) and on the initial parameters (pressure and temperature fields observed at a given time), to get a prediction of tomorrow's weather. In the case of *explanatory* models one relies on the match between predictions and empirical data to obtain potential explanations of a phenomenon of interest. One good example may be found in the explanation of the flocking behaviour observed for particular birds: although one does not know what is responsible for this emergent phenomenon, an agent-based simulation that ascribes particular behaviour rules to entities seems to reproduce the flocking behaviour; so a candidate explanation of the phenomenon is that birds, in nature, obey to rules that are similar to the ones implemented in the simulation (Huneman 2012).

As one can see the description of what *is* an explanatory or predictive model is highly intertwined with the issue of how the external world comes into play to constrain or test the model in question. I will thus first clarify the distinction between predictive and explanatory models and then argue that this distinction is the most relevant one if one wants to discuss the "testability" Kirchner seems to be interested in, and that the distinctions abstract/specific and what-if/actual are *not* the relevant categories in this discussion.

(i) Only explanatory models are testable₀.

Again, the reason why one builds an explanatory model is that one wants to identify the potential or candidate explanation for a particular phenomenon of interest. The aim is to identify the causal processes that can lead to a phenomenon. In order to do so, one builds a model that represents possible relations between properties (the model structure) before examining whether the model structure proposed is a good candidate explanation of the

phenomenon at stake. To know if it constitutes a candidate explanation, one has to test different putative scenarios and show that some lead to the prediction of the phenomenon of interest (these are the candidate explanations) while others do not. So clearly, here, testability_o (the comparison between the model predictions and empirical values) is a necessary procedure to discriminate between candidate explanations. So in the case of explanatory models, it is not only possible to carry out the comparison but also necessary given the epistemic aim of the model. Notice that the procedure cannot help one to *prove* that the candidate explanation is *the* actual explanation of the phenomenon; it will remain a candidate, or possible explanation only (Huneman 2012, Grüne-Yanoff and Weirich 2010). Along these lines, some have advocated the idea that the test of different scenarios constitutes a “numerical experiment” (Parker 2009; for contrasting views on the idea that numerical simulation allow one to carry out experiments, see Winsberg (2009), Grüne-Yanoff and Weirich (2010)).

This description concerns all explanatory models. Again, the explanatory/predictive distinction is orthogonal to the specific/abstract distinction. Testability is a straightforward process in the case of *specific* explanatory models (one just needs to compare the model predictions with empirical values obtained by other means); let us see how it would work for *abstract* explanatory models. I think it is in fact no more problematic: nothing prevents one from testing abstract explanatory models (in the sense of comparing the model predictions with observational data). Indeed nothing prevents one from abstracting away from empirical values and carrying out a qualitative comparison between these values and the model predictions. Kirchner is even caught red-handed doing it when he – pertinently – criticizes ecological Daisyworld model for predicting the stability of ecosystems when observations show the instability of such systems, (Kirchner 2003, 40).

(ii) Predictive models: why should we trust them?

The case of predictive models is different. Consider again meteorological models. Let us take for granted that they were built in order to have reliable predictions of tomorrow's weather. Needless to say, one wants the prediction before tomorrow evening. The event in question has not occurred yet (or, in other cases, one does not have access to it by any other means than through the model) so the output of the model cannot be compared with anything.

Now, regarding the possibility to compare the output with empirical values, there are two different categories of predictive models. In some cases, one will have a model that previously made predictions that one was able to compare with empirical values, and one now wants to use this model for a new (genuine) prediction for which one does not yet have empirical values. One may then inductively trust the new predictions because it happens that all the previous ones were successful, even if they were so in part for unknown reasons (some data assimilation procedures in meteorological models may be good examples of this category). In other cases, one *never* has been and *never* will be (in practice) able to compare *any* predictions of the model with empirical data. Either because the future is too distant, or because one will never have a proxy for the past phenomenon of interest. All GH models are of the latter category (the comparison is never possible). A good example is the lineage of "lifespan" models (which, interestingly, received the favour of Kirchner). The knowledge one expects to gain from these models is the amount of time during which Earth will remain habitable: this is an extreme case in which the prediction of the model will of course not be comparable, in theory or in practice, to the "actual value". So why should one rely on such predictive models?

Again, in the case of such predictive models one cannot compare the output with empirical data. It does not matter at all whether these predictive models are abstract or specific; their

predictive nature makes them not testable_o. The missing piece of the puzzle is the output, the prediction: one built the model *in order* to obtain the data the model predicts, because *one could not access these data otherwise*. There is no such option as “I don’t like models and prefer doing experiments” – (for a similar argument, see Winsberg 2003, 2006). This means that the reliability of the output of such predictive has to be ascertained on its own.

Let us try to clarify the situation. In the case of predictive models there is a system on which one thinks one has sufficiently good knowledge; one knows (or one relies on) the relation between the properties that are important to specify the dynamic of the system and one knows (or one relies on) the values of the parameters for a given state of the system. And what one wants to have is the description of the same system at a different time, in a different state. The initial state may be far back in the past or may be the actual state, and the final state (the prediction) may be in the future (e.g. meteorological models) or in the past (e.g. climate models of the precambrian): so “prediction” should not be understood here in its chronological sense, one can make predictions for past phenomena.

Here one relies on the predictions in virtue of relying on the source procedure that led to them: the initial conditions and parameters, the model structure, and the computational process. One may rely on the source procedure for different reasons: one used pieces of reliable knowledge (laws, theories, comparison with other models, etc.) that were established before or elsewhere, one used trustworthy model building techniques and made sure no relevant parameter was forgotten. When one wonders about the reliability of a model, one never wonders about the reliability of the “model as a whole” but always focuses on *parts* of the model, either the output, either the model structure: in the case of predictive model one wonders whether the model *predictions* are reliable (given your amount of trust in the model structure and the values of the parameters). By contrast, recall that for explanatory models one

examines whether the explanation proposed (the model structure) is a potential explanation (given the success of the comparison of the output with empirical values). So explanatory and predictive models are opposites as to their epistemic situation.⁸

Let us now examine how this would work when we introduce the specific/abstract categories, to show that they are not the relevant categories to discuss the testability of a model.

In predictive models one has to make sure that the modelling procedure accurately captures the properties of one's target system⁹. So there is a point where *discussion is open* as to whether the model faithfully represents the empirical world; it is here that the "external world" will constrain the model. And the striking point is that this "discussion" does not differ much in the case of abstract and in the case of specific models. Such a discussion occurred in the case of the lifespan models described above as an example of *specific* models. Several parameters were refined or added after Lovelock and Whitfield proposed their original model since these were shown to be quantitatively important: plant metabolism (Caldeira and Kasting 1992), a better parametrization of life's influence on weathering (Lenton and von

⁸ Notice that the same algorithm may alternatively be used to explain or to predict phenomena depending on the epistemic context in which it is used. Besides, I would be happy to grant that there are grey cases where it is not clear in which epistemic situation (prediction or explanation) we are.

⁹ Notice that one may rely on the source procedure even if one knows that *part* of the simulation artificially *misrepresents* the target but one has other reasons to believe in the validity of this model building technique (Winsberg 2006, Grüne-Yanoff and Weirich 2010).

Bloh 2001), and of geodynamic processes (Franck et al. 2000). But the very same sort of discussion can and did occur in the case of *abstract* models. Kirchner showed the problems that some hypotheses of Daisyworld create in misrepresenting life. He pointed out that the model does not allow the existence of black daisies whose optimal growth temperature would be higher than white daisies. Such a situation would create a pathological, instable Daisyworld (Kirchner 1989, 229). In other words the environmental preferences and the way in which the daisies affect their environment could not vary independently. The problematic hypotheses were henceforward abandoned (Wood et al. 2006, McDonald-Gibson et al. 2008, Wood et al. 2008). So for specific as well as for abstract models that claim to represent certain properties of life, one may always come up with counterexamples that demand adjustments of the model. The upshot for abstract models is that, although the *model* itself is a priori and is in a sense not meant to be testable, the *claim* that the model structure faithfully captures a particular empirical target is itself testable in the sense Kirchner demands (see Weisberg 2006 and Bedau 1999 for analogous points).

I would now argue that the epistemological status of the data produced by *what-if* and specific scenarios is the very same as the status of the data produced by predictive, specific and actual scenarios. Note that the unit of focus here is the scenario and not the model, since, within a what-if model it is more precisely the what-if scenario that may be thought to raise particular issues. One may indeed have specific concerns regarding the what-if scenario (and not regarding the actual scenario) in that its target does not (or cannot) exist out there (which is not the case for the actual scenario). But with respect to testability and to the reasons why one should trust a scenario, is there really a difference between the two? I think there is none. To come back to the example used in section 3.3 of this paper, one should rely on the “abiotic” curve (the *what-if* scenario) of Lenton and van Bloh’s lifespan model for the very same reasons that one relies on the biotic curve (the actual, specific and predictive scenario) in

figure 1.b. In both cases there is no possibility of comparing the prediction with empirical data. So in both cases, the reliability of the data comes from one's ability to make a good simulation based on sound source procedures. The fact that the target system genuinely exists or not is not relevant as to the status of the data produced. If the precedent claim is correct and if one does not worry about predictive and specific models, then one should not worry either about what-if scenarios (and thus not about what-if models).

The overall conclusion is that “abstract” and “what if” are not the right categories to discuss the procedures increasing the reliability of a model; the distinction between predictive and explanatory models is a better help. We have seen that only in the case of explanatory models can you compare the model predictions with empirical data. This should in no way lead us to disregard predictive models – which are pervasive in all sciences – but rather to acknowledge that the reasons why we rely on the predictions are different from the reasons why we rely on the model structure. Neither abstract nor what-if models should be thought to be outside of science, even when adopting an updated Popperian stance.

4.2 Interest of abstract and what-if models

Let us now turn to the reasons why one may be tempted to make abstract and what-if models. Why would one want to make such models? Why not be satisfied with specific and actual models after all? I would like to suggest three reasons:

- (i) Abstract and fictional models are good for how-possibly explanations.

We have seen that the entities involved in “abstract and fictional” models are meant to resemble actual organisms regarding *relevant abstract* properties (in the given context of investigation, properties that concern the interaction between organisms and their environment) but to differ from them regarding *irrelevant specific* properties (here the fact

that the entities are huge daisies). What can be the purpose of these apparently irrelevant *specific* details? In some cases they do not have any purpose but the model will be more specific than its target for contingent reasons or practical constraints.

I will however argue that in other cases – such as the original Daisyworld model – such specific details may serve a distinct purpose. It becomes more apparent if one realizes that the original model was meant to achieve a “how-possibly explanation” in the sense defined by Dray and discussed by Reydon (2012). How-possibly explanation in Dray’s sense are explanations that should not invoke laws of nature because how-possibly explanations do not intend to establish the necessity of the phenomenon to be explained. “As their aim is to establish that the occurrence of the explanandum is not entirely ruled out,” Reydon explains, how-possibly explanations “only need to specify those conditions that allow for the explanandum to occur even though given the theoretical and factual context it seemed unlikely or even impossible” (2012, 303).

The definition seems to fit perfectly with the aim of Daisyworld. Recall that it is because natural selection was presented as the sole explanation of organisms’ homeostasis and because of theoretical difficulties (absence of natural selection at a planetary level or problems raised by group selection) that the idea of a global environmental regulation was claimed to require planning or foresight by life. As Lovelock stated, Daisyworld was meant to address this critique:

The purpose of this article is to [...] present a simple model, called ‘Daisy World’, specifically developed to show that global homeostasis *does not need* foresight or planning by life. Lovelock 1983, 67, my emphasis.

In other words, Daisyworld was presented to show that even in the absence of natural selection at the global scale and without teleology involved, a regulation of the global

30

environment through the influence of life is not theoretically impossible. As pointed out by Kirchner (2002, 401), if the ambition of the paper was to claim that given our current knowledge of the abstract properties shared by all living beings, homeostasis or maintenance of habitability should *necessarily* emerge, then the model would not be able to meet this ambition since, as mentioned above, the original Daisyworld only illustrates a restricted case in which the environmental preferences and the way in which the daisies affect their environment can not vary independently. But if the ambition is only to show that homeostasis *could* emerge in plausible biological and geological contexts, then Daisyworld does the job (Kirchner 2002, 401).

This is precisely the reason why the “specific” details implemented in Daisyworld may ultimately matter: the model is supposed to exhibit a *detailed example* in which a phenomenon occurs whereas it was thought to be theoretically impossible. Although Reydon (2012, 309) seems to express scepticism about the existence of how-possibly explanations as defined by Dray, we have seen that the original Daisyworld may, interestingly, be a good candidate matching such a label. Weisberg (2013), after others (Brandon 1990), has claimed that “generalized modelling” (which corresponds to our “abstract models”) may be used to propose how-possibly explanations in what seems to be a broader acception of how-possibly explanations. Considering Reydon’s stricter conception leads us to claim that not *all* “abstract (or generalized) models” can be used to obtain how-possibly explanations: only “abstract and fictional” models can be used so, precisely because they contain more specific details than “abstract *simpliciter*” models.

(ii) Abstract *simpliciter* models are good for pursuing theorizing.

One reason to make abstract *simpliciter* models would be that abstract (computational) models are means (among others) of pursuing theorizing (Pigliucci 2013); such models offer a library

of candidate explanations for general phenomena. It is interesting in this respect to compare the hypotheses of Daisyworld models (see section 2.2 for examples) with what Scheiner and Willig (2008) think of as the basic statements of a general theory of ecology: “organisms are distributed in space and time in a heterogeneous manner”, “Organisms interact with their abiotic and biotic environments”, “Resources are finite and heterogeneous in space and time”, etc. In both activities, either theorizing with abstract models such as Daisyworld, or theorizing with verbal arguments in the case of Scheiner and Willig (2008), there is a focus on abstract, general properties shared by living organisms.

However, one may dislike theorizing. If there is something wrong with “theorizing in general” then the problem is not GH, but theorizing. So, though theorizing is a perfectly sound enterprise, it may have been plagued by misguided past attempts in GH: but one is then confronted with a scientific, and not an epistemological problem (some of the hypotheses were too strong). But if the problem is (mostly) scientific (which I think it is) then there is no epistemological problem inherent either to GH or to abstract and what-if models, in which case the two main points of my paper are made. Authors of the GH explain why they develop such abstract computational models as follows:

The computer may seem like an odd place to be looking for a planetary-sized phenomenon when we could be examining the real world. However, with a sample size of only one Earth the inferences that can be drawn about the likelihood of certain features are necessarily limited. (...) By creating many virtual worlds in the computer, we can begin to examine whether features we see on Earth, such as abundant recycling and environmental regulation, are likely or unlikely phenomena once life has emerged on a planet. (Lenton and Williams 2009, 61)

This echoes the debate in the GH literature about whether environmental variables on Earth have in fact been regulated by life and, if this is the case, whether it was due to *pure chance* or due to a *necessary* regulation by life (Watson 1999). Indeed, it may be the case that certain environmental variables are regulated on Earth, or that Earth would not have been habitable today if life had not influenced the environment. These questions can be investigated through concrete observations (oceanographic and atmospheric measurements) or thanks to *specific* computational models (the lifespan models). But in order to assess how likely (or necessary) the emergence of environmental regulation is (given the important properties of life), one needs to explore different counterfactual scenarios and one does so thanks to an *abstract* (computational) model. In order to evaluate the “likelihood of certain features”, which in the context of the above quotation is the likelihood of environmental regulation given the influence of life on its environment, one thus needs to assess the abstract properties of life. In this sense, the discussion over the legitimacy and epistemic purpose of GH abstract models echoes the debate in the A-life literature over the legitimacy of abstracting away to the fundamental properties of life while leaving aside what is taken to be contingent details, which questions the theoretical role that the concept of “life” may be able to play (Lange 1996, Bedau 2007).

Let me clarify further. When one wants to know how long Earth will remain habitable for *actual* living beings (the question tackled by the lifespan models), one only needs to know *what* is living (bacteria, photosynthetic organisms and complex organisms are), what the environmental variables that constrain their survival are, and how these will evolve in the future (notwithstanding the evolution of environmental preferences of organisms). However, GH authors seem to think that living beings have a special influence over their environment that makes them different from non-living entities. So they go further and put a special emphasis on what it means for an entity to be living in general. Let us consider the habitability

question now in an *abstract* mode: if one wants to show that living organisms, *qua* living beings, *necessarily* enhance habitability, then a correct explanation will require to show how some core features of what it means to be alive are linked with processes leading to an increase of the time during which the living entities at stake are maintained in their integrity (habitability). With abstract models GH postulates (or rather, it bets) that living beings share some properties that distinguish them from other entities in the world; as Lange (1996) would put it, they bet that life may be a natural kind and thus may become explanatory as such. On the questions raised by GH, recent abstract models have brought interesting results, constraining the conditions under which regulation can emerge given the nature of the relations between organisms and their environment: regulation cannot emerge if the characteristic timescale of the dynamic of organisms' adaptation to the environment is higher than the characteristic timescale of the dynamic of organisms' influence on their environment (Williams and Noble, 2005, Weaver and Dyke, 2012), natural selection acting at the level of the community can enable the emergence of a regulation at the level of the ensemble of the communities (Williams and Lenton, 2008, see Wood et al. 2008, Dutreuil 2013 for a review of the other principal results brought by the Daisyworld literature).

So GH has developed abstract models to abstract away from contingent details about life and its environment and highlight a number of core features, just as A-life does. Abstract (computational) models may be useful when evaluating the likelihood of the clustering of certain properties while empirical inquiries will be confronted with a unique sample.¹⁰ This is something specific and actual models simply *cannot* do; so the issue lies not with choosing a

¹⁰ Huneman, this issue, offers a detailed and interesting discussion of the way the evaluation of the likelihood can be carried out.

strategy for model building but with accepting the question at stake as a legitimate question or not: is it legitimate to bet on the contingency of certain features of life? It may turn out not to be so, but the debate is much larger than GH *per se* and pertains to theoretical biology, A-life and the philosophy of biology at large.

(iii) What-if models are important to attribute causal or explanatory power.

After having discussed the interest of abstract models, let us examine the role played by what-if models. Please, notice that the unit of analysis is not the what-if *scenario* but the what-if *model* (comprising the what-if scenario and the “normal” or “actual” scenario). This modelling practice, *qua* modelling practice, seems to have received less attention from philosophers of science than, say, “abstract (mathematical) modelling”, whereas they are clearly distinct. What-if models have been discussed by Weisberg (2013) when he examined Fisher’s strategy to explain *why biological populations have only two sexes*. In order to do so, Fisher imagines a three sex population with a formal model and explores the consequences of such an hypothesis. At the end of the day, what one expects from these models is not so much the description of a world that is possible but not actual. One temporarily needs this description in order to compare it with the actual scenario and to conclude that the difference between the two (actual and what-if scenarios) can be imputed to the variable artificially switched off or changed.

One may thus compare worlds with and without the influence of life so as to bring into relief the role played by life in the actual world. What-if models may here be crucial steps in the *explanation* of a phenomenon. This interpretation would fit with the conception of causality defended by Woodward (2010) and seems to fit perfectly with the account of explanation given by Strevens (2004, ch. 4.3). The procedure of progressively removing parameters from a model (the “eliminative procedure” described by Strevens) to see if the phenomenon to be

explained still holds after this removal is a crucial step in determining whether the parameter is a difference-maker and whether it is a good explanation of the phenomenon of interest. Although the space for developing the argument is lacking here, it seems that this strategy (“what-if” or eliminative procedure) would be as essential to non-causal accounts of explanation as it is to Strevens’ causal account. Let us see how this would work with Huneman’s (2010) topological explanations, a kind of explanation that may be thought of as non-causal (but need not be so depending on one’s conception of causality): artificially removing connections or knots from a topological networks to see if certain properties (robustness, stability, etc.) still hold seems to be an important step in attributing explanatory power to the given topology of the initial network.

Let me summarize the role of the different kinds of model discussed in the preceding:

- “Abstract and fictional” models may enable one to obtain how-possibly explanations (they help one to show that an event that was previously thought to be theoretically impossible is in fact possible); the original Daisyworld of Watson and Lovelock was used in this way.
- “Abstract *simpliciter*” models allow one to pursue theorizing and explore the likelihood of the clustering of certain properties; some later variants of Daisyworld as well as A-life models such as FLASK were used for these purposes, to assess the conditions under which “regulation” will emerge.
- “What-if” models may be construed as a necessary enterprise in imputing explanatory (or causal) power to one variable of interest. Different models were presented to use this strategy: Daisyworld, Lenton and Von Bloh lifespan model.

In the section dedicated to the description of the models we raised the question of the epistemic priority of “abstract *simpliciter*” over “abstract and fictional” models: in the past decade, some Daisyworld models were explicitly made “abstract *simpliciter*”, whereas others, such as Boyle et al. (2011) or Weaver and Dyke (2012), fully aware of Kirchner’s critiques, conserved the “abstract and fictional style”. The issue of giving priority to one or the other model only makes sense in the context where they are used to obtain similar knowledge. Boyle et al. (2011), as well as Weaver and Dyke (2012), did not propose “abstract and fictional” models to obtain how-possibly explanation, in contrast to Watson and Lovelock (1983). They rather pursued the kind of knowledge obtained by “abstract *simpliciter*” models. Is one strategy better than the other? I would argue that they are epistemologically *on a par*, although one may have an aesthetic preference for one or the other. Indeed one can wonder what is the *relevant* difference (in the context where one wants to obtain abstract knowledge) between Daisyworld style models and abstract models such as FLASK. There is none. The first model describes fictive specific entities with which we are not familiar (huge daisies) while the second describes more familiar entities (flasks of microbes), but then both warn the reader: the specific story does not matter, so please, abstract away from these details (do not look for giant daisies in the world, do not look for specific nutrients). Since the specific details do not matter given the purpose of the model, then whether these details are “fictive” or more “familiar” should not worry us.

This brings me back to the idea implicitly developed by Kirchner (and recalled in the second section) that some models may be of greater value than others. We have seen here that the different models (specific, abstract, actual, what if) serve different epistemic purposes and help us answer different kinds of question. They are not different ways of pursuing a common goal, but different ways of pursuing different goals: how-possibly explanations for actual phenomena (abstract and fictional), evaluation of the likelihood of the clustering of some

properties (abstract *simpliciter*), attribution of causal or explanatory power to one variable of interest (what-if). I have argued that to engage oneself in trying to explore any of these directions is an epistemologically valid enterprise. If it turns out that all different models (specific, abstract, what-if) are epistemologically sound then the debate over the relative interest of certain models vs. others (in Kirchner's terms, whether there are some more "pressing" issues than others) can not be discussed on a purely epistemological ground. The issue is not how to best allocate epistemic resources to attain a unique goal, it is a question of comparing different epistemologically legitimate goals. So the only pertinent level of argumentation is social and not epistemological.

5 Conclusion

Let us sum up the main arguments developed throughout this paper.

First, it is simply not the case that testability – understood as the comparison of the output of a model to empirical data – is always a necessary procedure to have reliable models. This procedure is only available in the case of explanatory models when one wants to discriminate between possible explanations implemented in the model structure (but even in these cases, this procedure is not the only one important for assessing the reliability of the model). In the case of predictive models, the comparison is not possible. One then uses the model *in order* to obtain the output (the prediction) and one relies on the prediction in virtue of relying on the source procedures that enabled the predictions (the initial parameters and the model structure). The prominent role that the explanatory/predictive distinction plays when it comes to the reliability and testability of a model makes it clear that the initial typology based on Kirchner's critique (actual and specific, abstract *simpliciter*, abstract and fictional, what if models) is of little help in this context.

However – and this is the second principal argument of the paper – this typology is useful when it comes to discussing other epistemic purposes of the models. We have seen that the discussion over the (relative) interest of these different models should not be cashed out in terms of strategy since these different kinds of model are used to pursue different epistemic roles: how-possibly explanations in the case of abstract and fictional models, pursuing theorizing and exploring the likelihood of the clustering of certain properties in the case of abstract *simpliciter* models, attribution of causal or explanatory power to one variable of interest in the case of what-if models.

These two main arguments, although stemming from a discussion of a given field of science (GH), aimed at contributing to general philosophical considerations on the role of models. Back to GH now, they rebut the principal epistemological criticisms addressed to GH models by Kirchner. Besides, this analysis sheds light on one particular aspect of GH: the special emphasis that it puts on the role of life. Consider first the particularity of the what-if models of GH in which the influence of life is switched off. The authors never switch off, say, the influence of geodynamics on the climate, but always contrast scenarios with and without the influence of life. This brings into relief the particular interest GH shows for life: GH is above all an hypothesis about the *consequences* of life on Earth. Consider now the implication of developing abstract models of the interaction of life with its environment. These models require something that actual and specific models do not: they require a *definition of life*.

The postulate that living beings may be distinct from non-living entities is in fact a bet; it may very well in the end turn out that the influence of life has nothing special as to its magnitude, as to how it affects habitability or how it regulates the environment. Interestingly, GH, by working on such questions and deploying such strategies to answer them, joins up with the fields of research (A-life, exobiology, origins of life) which, given their question of interest,

need to rethink the problem of the definition of life (something the other branches of biology need not do; see Morange 2003, Gayon 2010). This should not come as a surprise since the definition of life was one of the fundamental starting points of GH (Lovelock 1965) and since the fields of research mentioned above emerged from roughly the same institutional and intellectual context (Dick and Strick 2005).

Acknowledgement

I am indebted to all those present at the EASPLS meeting (Geneva, September 2012) and at the Duke Consortium for the history and philosophy of biology (June, 2013) for giving me their comments on these ideas, especially to Marie Kaiser and Tyler Curtain. I would like also to thank warmly Frédéric Bouchard, Jean Gayon, Philippe Huneman, Arnon Levy, Arnaud Pocheville and Judith Villez as well as an anonymous reviewer who helped me significantly improve earlier versions of this paper. I must finally thank Staffan Müller-Wille for his careful reading which has greatly contributed to the clarification and amelioration of the submitted version.

References

Barberousse A. and Vorms M., 2013, “Computer simulations and empirical data”. In: Durán J.M. and Arnold E. (Eds.), *Computer simulations and the changing face of scientific experimentation*. Cambridge Scholars publishing, 29–45.

Bedau M., 2007, “Artificial Life”. In: Gabbay W., Thagard P., Woods J. (Eds.), *Handbook of the philosophy of science, Volume 3: Philosophy of biology*. North Holland: Elsevier, 595–613.

- Bedau M.A., 1999, "Can unrealistic computer models illuminate theoretical biology". In: Annie S. Wu (Ed.), *Proceedings of the 1999 genetic and evolutionary computation conference workshop program*, Orlando, Florida, July 13, 20–23.
- Boyle R., Lenton T. and Watson A., 2011, "Symbiotic physiology promotes homeostasis in daisyworld", *Journal of theoretical biology*, 274(1): 170–182.
- Brandon R., 1990, *Adaptation and environment*. Princeton, NJ: Princeton University Press.
- Caldeira K. and Kasting J., 1992, "The life span of the biosphere revisited", *Nature*, 360(6406): 721–723.
- Dawkins R., 1982, *The extended phenotype: The gene as the unit of selection*. Oxford: Oxford University Press, USA.
- Dick S.J. and Strick J. E., 2005, *The living Universe: NASA and the development of astrobiology*. New Brunswick, New Jersey, and London: Rutgers University Press.
- Dietrich W. and Perron J., 2006, "The search for a topographic signature of life", *Nature*, 439(7075): 411–418.
- Doolittle W., 1981, "Is nature really motherly", *CoEvolution Quarterly*, 29: 58–63.
- Downing K. and Zvirinsky P., 1999, "The simulated evolution of biochemical guilds: reconciling Gaia theory and natural selection", *Artificial Life*, 5(4): 291–318.
- Dutreuil S., 2012, "L'hypothèse Gaïa : pourquoi s'y intéresser même si l'on pense que la Terre n'est pas un organisme?", *Bulletin de la société d'histoire et d'épistémologie des sciences de la vie*, 19(2): 229–241.

Dutreuil S., 2013, "Comment le modèle Daisyworld peut-il contribuer à l'hypothèse Gaïa?". In : Varenne F. and Silberstein M. (Eds.), *Modéliser et simuler. Epistémologies et pratiques de la modélisation et de la simulation. Tome 1*. Paris: Editions matériologiques.

Dyke J. and Harvey I., 2006, "Pushing up the daisies". In: Rocha L.M. (Ed), *Artificial Life X, proceedings of the tenth international conference on the simulation and synthesis of living systems*, Cambridge: MIT Press, 426–431.

Dyke J. and Weaver I., 2013, "The emergence of environmental homeostasis in complex ecosystems", *PLoS computational biology*, 9(5).

Forber P., 2010, "Confirmation and explaining how possible", *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 41(1): 32–40.

Franck S., Block A., VonBloh W. et al., 2000, "Reduction of biosphere life span as a consequence of geodynamics", *Tellus B*, 52(1): 94–107.

Franck S., Bounama C. and VonBloh W., 2006, "Causes and timing of future biosphere extinctions", *Biogeosciences*, 3(1): 85–92.

Free A. and Barton N., 2007, "Do evolution and ecology need the Gaia hypothesis?", *Trends in Ecology and Evolution*, 22(11): 611–619.

Frigg R., 2010, "Models and fiction", *Synthese*, 172(2): 251–268.

Gayon J., 2010, "Defining life: synthesis and conclusions", *Origins of Life and Evolution of Biospheres*, 40(2): 231–244.

Godfrey-Smith P., 2006, “The strategy of model-based science”, *Biology and Philosophy*, 21(5): 725–740.

Godfrey-Smith P., 2009, “Models and fictions in science”, *Philosophical Studies*, 143(1): 101–116.

Grüne-Yanoff T. and Weirich P., 2010, “The philosophy and epistemology of simulation: A review”, *Simulation and Gaming*, 41(1): 20–50.

Harvey I., 2004, “Homeostasis and rein control: From daisyworld to active perception”. In: Pollack J. et al. (Eds), *Proceedings of the ninth international conference on the simulation and synthesis of living systems, ALIFE, 9*, Cambridge: MIT Press,. 309–314.

Humphreys P., 2009, “The philosophical novelty of computer simulation methods”, *Synthese*, 169(3): 615–626.

Huneman P., 2010, “Topological explanations and robustness in biological sciences”, *Synthese*, 177(2): 213–245.

Huneman P., 2012, “Computer science meets evolutionary biology: Pure possible processes and the issue of gradualism”. In: Pombo O. et al. (Eds), *Special Sciences and the Unity of Science*, Springer, 137–162.

Kirchner J., 1989, “The Gaia hypothesis: can it be tested”, *Review of Geophysics*, 27(2): 223–235.

Kirchner J., 2002, “The Gaia hypothesis: Fact, theory, and wishful thinking”, *Climatic Change*, 52(4): 391–408.

Kirchner J., 2003, 'The Gaia hypothesis: conjectures and refutations' *Climatic Change*, 58(1): 21–45.

Lakatos I., 1978, "Science and pseudo science". In: Worrall J. and Currie G., *The Methodology of Scientific Research Programmes, Philosophical papers, Volume 1*, Cambridge: Cambridge University Press, 1–7.

Lange M., 1996, "Life, 'artificial life', and scientific explanation", *Philosophy of Science*, 63(2): 225–244.

Lenton T. and Von Bloh W., 2001, "Biotic feedback extends the life span of the biosphere", *Geophysical Research Letters*, 28(9): 1715–1718.

Lenton T. and Williams H., 2009, "Gaia and Evolution". In: Crist E. and Rinker H.B., *Gaia in Turmoil: Climate Change, Biodepletion, and Earth Ethics in an Age of Crisis*, Cambridge: MIT Press, 61–85.

Levy A., 2012, "Models, fictions and realism: Two packages", *Philosophy of Science*, 79(5): 738–748.

Li K., Pahlevan K., Kirschvink J., and Yung Y., 2009, "Atmospheric pressure as a natural climate regulator for a terrestrial planet with a biosphere", *Proceedings of the National Academy of Sciences*, 106(24), 9576–9579.

Lovelock J., 1983, "Daisy world – a cybernetic proof of the gaia hypothesis", *Coevolution Quarterly*, Summer 1983(38): 66–72.

Lovelock J. and Margulis L., 1974, "Atmospheric homeostasis by and for the biosphere: the Gaia hypothesis", *Tellus*, 26(1-2): 2–10.

- Lovelock J. and Whitfield M., 1982, “Life span of the biosphere”, *Nature*, 296: 561–563.
- Lovelock J., 1965, “A physical basis for life detection experiments”, *Nature*, 207: 568–570.
- McCain K. and Weslake B., 2013, “Evolutionary theory and the epistemology of science”. In: Kampourakis K. (Ed.), *The philosophy of biology: a companion for educators*. Springer.
- McDonald-Gibson J., Dyke J., DiPaolo E. and Harvey I., 2008, “Environmental regulation can arise under minimal assumptions”, *Journal of theoretical biology*, 251(4): 653–666.
- Morange M., 2003, *La Vie expliquée? : 50 ans après la double hélice*. Paris: Odile Jacob.
- Oreskes N., Shrader-Frechette K., Belitz K. et al., 1994, “Verification, validation, and confirmation of numerical models in the earth sciences”, *Science*, 263(5147): 641–646.
- Parker W.S., 2009, “Does matter really matter? Computer simulations, experiments, and materiality”, *Synthese*, 169(3): 483–496.
- Pigliucci M., 2008, “The borderlands between science and philosophy: an introduction”, *The Quarterly Review of Biology*, 83(1): 7–15.
- Pigliucci M., 2013, “On the different ways of ‘doing theory’ in biology”, *Biological Theory*, 7(4): 287–297.
- Pocheville A., 2010, *La niche écologique: concepts, modèles, applications*. PhD thesis, Ecole Normale Supérieure, Ecole doctorale Frontières du Vivant.
- Reydon T.A., 2012, “How-possibly explanations as genuine explanations and helpful heuristics: A comment on Forber”, *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(1): 302–310.

Ruse M., 2013, *The Gaia hypothesis: science on a pagan planet*. Chicago: The University of Chicago Press.

Sarkar S. and Plutynski A., 2010, *A Companion to the philosophy of biology*. USA: Blackwell Publishing Ltd.

Scheiner S.M. and Willig M., 2008, “A general theory of ecology”, *Theoretical Ecology*, 1(1): 21–28.

Sober E., 1991, “Learning from functionalism: prospects for strong artificial life”. In: Langton C., Taylor C., Farmer J.D. and Rasmussen S. (Eds.), *Artificial Life II*. Boulder: Oxford WestviewPress, 749–766.

Sober E., 2008, *Evidence and evolution: the logic behind the science*. USA: Cambridge University Press.

Strevens M., 2004, “The causal and unification approaches to explanation unified – causally”, *Noûs*, 38(1): 154–176.

Watson A., 1999, “Coevolution of the Earth’s environment and life: Goldilocks, Gaia and the anthropic principle”. In: Craig G.Y., Hull J.H. (Eds), *James Hutton – Present and Future*. Geological Society, London, Special Publications, 150, 75–88.

Watson A. and Lovelock J., 1983, “Biological homeostasis of the global environment: the parable of Daisyworld”, *Tellus B*, 35(4): 284–289.

Weaver I. and Dyke J., 2012, “The importance of timescales for the emergence of environmental self-regulation”, *Journal of Theoretical Biology*, 313(21): 172–180.

Weisberg M., 2006, “Robustness analysis”, *Philosophy of Science*, 73(5): 730–742.

Weisberg, M., 2007, “Who is a modeler?”, *The British journal for the philosophy of science*, 58(2): 207–233.

Weisberg M., 2013, *Simulation and similarity: using models to understand the world*, USA: Oxford University Press.

Williams H. and Lenton T., 2008, “Environmental regulation in a network of simulated microbial ecosystems”, *Proceedings of the National Academy of Sciences*, 105(30), 10432–10437.

Williams H. and Noble J., 2005, “Evolution and the regulation of environmental variables”. In: Capcarrere M.S. et al., *Proceedings of VIIIth conference on Artificial Life*, Berlin: Springer, 332–342.

Winsberg E., 2003, “Simulated experiments: Methodology for a virtual world” *Philosophy of Science*, 70(1): 105–125.

Winsberg E., 2006, “Models of success versus the success of models: Reliability without truth”, *Synthese*, 152(1): 1–19.

Winsberg E., 2009, “A tale of two methods”, *Synthese*, 169(3): 575–592.

Wood A., Ackland G. and Lenton T., 2006, “Mutation of albedo and growth reponse produced oscillations in a spatial Daisyworld”, *Journal of Theoretical Biology*, 242: 188–198.

Wood A., Ackland G., Dyke J., Williams H. and Lenton T., 2008, “Daisyworld: A review”, *Reviews of Geophysics*, 46(1).

Woodward J., 2010, “Causation in biology: stability, specificity, and the choice of levels of explanation”, *Biology and Philosophy*, 25(3): 287–318.