2-2017

Talk about toy models

Joshua LUCZAK
Singapore Management University, jluczak@smu.edu.sg

Talk About Toy Models

Joshua Luczak
Leibniz Universität Hannover
joshmluczak@gmail.com

Forthcoming in *Studies in History and Philosophy of Modern Physics*

Abstract

Scientific models are frequently discussed in philosophy of science. A great deal of the discussion is centred on approximation, idealisation, and on how these models achieve their representational function. Despite the importance, distinct nature, and high presence of toy models, they have received little attention from philosophers. This paper hopes to remedy this situation. It aims to elevate the status of toy models: by distinguishing them from approximations and idealisations, by highlighting and elaborating on several ways the Kac ring, a simple statistical mechanical model, is used as a toy model, and by explaining why toy models can be used to successfully carry out important work without performing a representational function.

Introduction

Scientific models are frequently discussed in philosophy of science. A great deal of the discussion is centred on approximation, idealisation, and on how these models achieve their representational function. Some philosophers have tried to get clear on approximations and idealisations.[1] Some dissect them, typically idealisation, into types.[2] Some have highlighted and discussed their differences.[3] Some have discussed their representational functions and ideals.[4] Many have discussed how these models achieve their representational function.[5] Some have questioned and discussed what these models can tell us about reality.[6] Some have questioned and discussed their role in scientific explanations.[7] The list goes on.

Toy models are easily confused with approximations and idealisations. They are, however, distinct. Unlike idealisations and approximations, toy models do not perform a representational function. That is, they do not represent anything. They are nonetheless important for science. Toy models are models that are not intended to perform a representational function, but rather to perform one or more of the following functions:

1. To learn to use, or to become comfortable with, certain formal techniques (e.g. renormalization). That is, as a pedagogical device.

2. To elucidate certain ideas relevant to a theory. That is, to reach a clearer understanding of an idea, its implications, and its relation to other ideas within a theory.

3. To test the compatibility of various concepts (i.e. in a consistency proof).

4. To generate hypotheses about other systems.

One commonly finds authors using simple models to perform one or more of these functions in the introductory chapters of physics textbooks. When they do so, and do not intend for them to perform a representational function, it is appropriate to regard their models as toy models. Some examples common to statistical mechanics include: Mark Kac's ring model, Paul and Tatyana Ehrenfests' urn (dog-flea) and wind-tree models, the baker's transformation, the Ising model, and the Arnold cat map.

[1] See, for example, Achinstein (1968), Bunge (1973), and Laymon (1990).

[2] See, for example, Frigg and Hartmann (2012: Sec.1.1), McMullin (1985), and Weisberg (2007).

[3] See Norton (2012), for example.

[4] See Weisberg (2007), for example.

[5] See, for example, Aronson, Harré, and Way (1995), Bailer-Jones (2003), Bartels (2006), Frigg and Hartmann (2012) (and the references therein), Giere (2010), Godfrey-Smith (2006), Mundy (1986), Pincock (2012), Suárez (2015), Suppes (2002), Swoyer (1991), and Weisberg (2013).

[6] See Frigg and Hartmann (2012: Sec.3) and Swoyer (1991), for example.

[7] See, for example, Bokulich (2009, 2011, 2012), Cartwright (1983: Ch. 8), Elgin and Sober (2002), Frigg and Hartmann (2012: Sec. 5.4), and Woodward (2003).

Despite the importance, distinct nature, and presence of toy models, they have received little attention from philosophers.[8] Perhaps this is because many philosophers and scientists mistakenly think of them as either a kind of approximation or idealisation. Or perhaps it is because many of the models used in the ways listed above are, on other occasions, used to represent other systems, and this obscures the distinction between these model types. Whatever the case may be, it would be beneficial for our understanding of scientific models to more deeply explore toy models and to engage in focused discussions of them. This paper hopes to advance this goal. It aims to elevate the status of toy models and to encourage more focused discussions of them. This will be achieved by distinguishing them from approximations and idealisations, by highlighting and elaborating on several ways the Kac ring is used as a toy model, and by explaining why it can be successfully used in these ways without performing a representational function. This paper will focus on ways in which the Kac ring can be used to successfully perform functions **2-4**, since the claim that it can be used to successfully achieve **1**, without performing a representational function, is uncontroversial. In speaking by way of the Kac ring model, this paper intends to support the claim that toy models play an important role in science, despite them not performing a representational function.

The next section notes some standard claims made about approximations, idealisations, and scientific representation. These are noted so as to distinguish approximations and idealisations from toy models. Parts of the third section draw on the work of Gottwald and Oliver (2009: Sec.3). The section begins by highlighting some of the Kac ring's features. The model is then used to elucidate two important statistical mechanical ideas: the reversibility objection and the recurrence objection. Its ability to successfully perform this task without performing a representational function is also discussed in this section. Discussion of the recurrence and reversibility objections encourages using the Kac ring in other ways: as part of a consistency proof and to suggest interesting things about other systems, including real systems. The fourth section includes the consistency proof and a discussion of why an agent can use the model to successfully perform these tasks, despite it not performing a representational function. The fifth section highlights and continues to discuss its use in generating hypotheses about other systems without performing a representational function. The paper ends with a few concluding remarks and with some suggestions about the direction of future work.

Approximations, Idealisations, Scientific Representations, and Toy Models

Approximations and idealisations are frequently discussed in the literature on modelling. Despite there being disagreement about how to precisely characterise these model types, there are certainly some things that can be said about them that are uncontroversial. An approximate

[8] Michael Weisberg (2013: Ch.7.3) has discussed, what he calls, targetless models. Targetless models are like toy models in that they are not intended to perform a representational function. It is unclear, however, from Weisberg's short discussion of targetless models, whether they are similar to toy models in other respects—such as how and why they can be used to carry out important work.

model inexactly represents a target system. Typically this is some aspect of the world. Idealised models also represent target systems. These models, however, introduce some kind of deliberate simplification or distortion. These modifications are usually introduced to make it easier to deal with the target. Importantly, approximations and idealisations perform a representational function.

While it is a subject of debate within philosophy of science as to what exactly constitutes a model's representation of a target, there are good reasons to think that the representation can neither simply be reduced to a similarity relation that holds between the model and the target nor to some kind of morphism relation (e.g. isomorphism) that holds between the structures that are instantiated by both the model and the target.[9] On the positive side, it seems fair to say that a model performs a representational function only if its user intends for it to perform a representational function. This fits with a promising and growing view that scientific representation is a practice performed by intentional agents.[10] As Ronald Giere (2004: p.747) explains, scientists use models to represent portions of the world for various purposes. But, as he continues, it is not the model that is doing the representing; it is the scientist using the model who is doing the representing. If we embrace both of these views of scientific representation, as we will in the remainder of this paper, then we can maintain that models can be used as toy models and that these models do not perform a representational function. That is, that an agent can use a model to successfully perform one or more of the functions **1-4** without intending that it perform a representational function and that in these circumstances these models do not perform a representational function. The next few sections intend to highlight these facts, and to explain why they are so, for functions **2-4**.

Models, such as the Kac ring, can be used by agents to do a lot of interesting and important work simply because either they instantiate certain properties (see the discussions of functions **2-3**) or because they instantiate certain properties that are also known to be instantiated by other systems (see the discussion of function **4**). In the latter type of case, these similarities permit treating the model as an analogue. Moreover, these similarities permit, and are sufficient for, analogical reasoning.[11] That is, they permit, and are sufficient for, employing some version of the following argument schema, where S is some model and T is some other system:

P1. S is similar to T in certain (known) respects.

P2. S has some further feature Q.

C. Therefore, T also has the feature Q, or some feature Q^* similar to Q.

[9]See Suárez (2015) and Suárez (2003) for more on this point. And see Suárez (2015) for a state-of-the-art review of the philosophical literature on scientific representation.

[10]See Suárez (2015).

[11]See Bartha (2013) for a comprehensive discussion of analogies and analogical reasoning.

Importantly, however, as Mauricio Suárez (2015) has argued, by drawing on Nelson Goodman's (1968) argument against resemblance theories of artistic representation, similarity is not sufficient for representation.[12] Similarity, as it is typically understood, is a reflexive and symmetric relation. Scientific representations, on the other hand, do not have these logical properties. If similarity were sufficient for representation then, for example, a dilute gas would be a scientific representation of a billiard ball model and itself. But it is neither of these things, so similarity is not sufficient for scientific representation. So then, even if a model instantiates properties that are also instantiated by other systems, this does not entail that it represents any or all of those systems, or anything at all.

The Kac Ring

The Kac ring first appeared in a series of lectures given by Mark Kac in1959 at the University of Colorado. The purpose of these lectures was to furnish an introduction to probability theory and its applications to an audience that had little knowledge of these subjects.[13] In a lecture on classical statistical mechanics, Kac (1959: p.99) used the ring model to introduce his audience to the statistical mechanical treatment of irreversible phenomena.

The Kac ring is a simple, explicitly solvable model. It is similar, in certain respects, to a dilute gas. In the model, N *sites* are arranged around a circle, forming a one-dimensional *periodic lattice*. Sites are joined to their neighbours by an *edge*, and $0 < n < N$ of the edges carry a marker. Each site is occupied by either a black ball or a white ball.[14] The balls and markers are similar to the molecules that comprise a gas.

The system evolves on a discrete set of ticks $t \in \mathbb{Z}$ from state t to state $t+1$ in the following way: each ball moves in a clockwise direction to its nearest neighbour. When a ball passes a marker its colour changes. This is analogous to changes in the velocities of molecules of a gas as they collide with one another.

The Kac ring has a number of interesting features. Its microdynamics are symmetric under time-reversal. Any reversed sequence of states is compatible with the dynamics of the system. The system is strictly periodic, and so displays recurrence. The ring, after a series of ticks, returns to its initial state. After N ticks, each ball has reached its initial site and changed colour n times. If n is even, the initial state recurs. If n is odd, it takes at most $2N$ ticks for the initial state to recur.

The Kac ring can be used, as it often is, to elucidate ideas relevant to statistical mechanics and thermodynamics.[15] A nice example is its use in elucidating the reversibility and recurrence objections; two important statistical mechanical concepts.

[12] See also Suárez (2003).

[13] See the forward and preface to Kac (1959).

[14] Similar descriptions of the model can be found in Bricmont (1995: Appendix 1), Bricmont (2001: p.10), Dorfman (1999: Sec.2.3), Kac (1959: p.99), and Gottwald and Oliver (2009: Sec.3).

[15] See, for example, Bricmont (1995: Appendix 1), Bricmont (2001), Dorfman (1999: Sec.2.3), Kac (1959: Ch.3 Sec.14-15), Gottwald and Oliver (2009), Schulman (1997: Sec.2.1), and Thompson (1972: Sec.1.9).

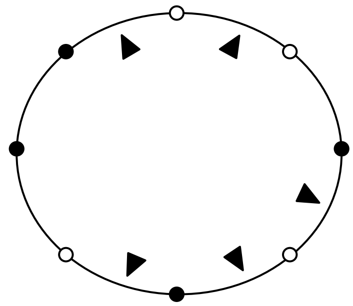

Figure 1: A Kac ring with $N = 8$ lattice sites and $n = 5$ markers.

We say that an isolated macroscopic system is behaving thermodynamically if it is in equilibrium or if it is spontaneously approaching equilibrium. Every macroscopic system we encounter seems to exhibit thermodynamic behaviour. It is a goal of classical statistical mechanics to account for this behaviour in terms of the behaviour of the microscopic parts that constitute these systems.[16] These systems are standardly described as having a microdynamics that is symmetric under time-reversal. They also display recurrence, if they have bounded phase space energy hyper-surfaces.

Ludwig Boltzmann famously attempted to account for thermodynamic behaviour in a classical framework. Or, at least, one instance of it. In 1872, Boltzmann considered how the distribution of velocities of the molecules of a contained dilute gas could be expected to change under collisions and argued that there was a unique distribution—now called the Maxwell-Boltzmann distribution—that was stable under collisions.[17] Boltzmann further argued that a gas that initially had a different distribution would move toward the Maxwell-Boltzmann distribution. To argue for this, Boltzmann defined a quantity, which we now call H, showed that it reached a minimum value for the Maxwell-Boltzmann distribution, and argued that it would *monotonically* decrease to its minimum.[18] This result is now known as Boltzmann's H-theorem. It is a straightforward consequence of Boltzmann's transport equation. Importantly, it is a temporally asymmetric result.[19]

[16] It is worth noting that this is not the only or primary goal of statistical mechanics. See Wallace (2013) for more on this point.

[17] See Boltzmann (1872).

[18] The quantity we call H was originally denoted E in Boltzmann's early work. See Boltzmann (1872).

[19] See Brown, Myrvold, and Uffink (2009) for more on Boltzmann's H-theorem.

In the wake of this result many began to wonder how Boltzmann arrived at it, having only assumed a dynamics that is symmetric under time-reversal. It was later discovered that he did not, and two famous objections have shown that he could not. These are known as the reversibility and recurrence objections. The former is usually credited to Josef Loschmidt and the latter to Ernst Zermelo.[20] The reversibility objection applies to systems whose microdynamics are symmetric under time-reversal. In the case of Boltzmann's gas, it says that for any set of trajectories of the molecules of the gas, the time-reversed trajectories are also compatible with the dynamics. So not *all* microstates of the gas at *any* time lead to a *monotonic* decrease of H. The recurrence objection applies to classical systems with bounded phase space energy hypersurfaces. That is, to systems, with total fixed energy, such as Boltzmann's gas. If we consider a small open neighbourhood of the system's initial state, and ask, will the system, after it leaves that neighbourhood, ever return to it? Then the answer, which makes use of Henri Poincarè's recurrence theorem, is *yes*, it will, for almost all initial phase space-points, i.e. for all except a set of Lebesgue measure zero. More plainly, but less precisely, the objection notes that no initial microstate will yield a *monotonic* decrease of H.

To derive Boltzmann's original, asymmetric, result, one needs more than what is given by simply applying Newton's laws of motion to molecular collisions. For Boltzmann, it was a temporally asymmetric assumption that appeared in the derivation of his transport equation. The assumption, which posits an absence of correlations between the velocities of colliding molecules at all times, is now known as the *Stoßzahlansatz*.[21]

This brief discussion has touched on two interesting and important foundational statistical mechanical concepts. These ideas are also important, historically. They may, however, be difficult to grasp or to fully appreciate—especially for those new to statistical mechanics. Happily, many are able to reach a clearer understanding of the objections, their implications, and how they relate to other ideas in statistical mechanics, by examining the Kac ring.

Recall the model. Let $B(t)$ denote the total number of black balls and $b(t)$ the number of black balls that pass a marker on the next tick. Similarly, let $W(t)$ denote the number of white balls and $w(t)$ the number of white balls that pass a marker on the next tick. It follows that

$$B(t+1) = B(t) + w(t) - b(t) \tag{1}$$

and

$$W(t+1) = W(t) + b(t) - w(t). \tag{2}$$

We can study the difference between the number of black and white balls at various times.

[20] See Uffink (2007) and Brown et al. (2009) for more on these objections.

[21] The term "*Stoßzahlansatz*" was coined by Paul and Tatiana Ehrenfest (1907). See Uffink (2007) and Brown et al. (2009) for more on the *Stoßzahlansatz*.

$$\Delta(t) = W(t) - B(t) \tag{3}$$

and

$$\Delta(t+1) = W(t+1) - B(t+1) = \Delta(t) + 2b(t) - 2w(t). \tag{4}$$

The system is in equilibrium when $W(t) \approx B(t)$. W, B, and Δ are *macroscopic* quantities. Many different microstates give rise to the same macroscopic quantities. In contrast, w and b give local information about individual sites. They cannot be calculated without knowing the location of each marker and the colour of the ball at each site. Importantly, the *evolution* of W, B, and Δ cannot be determined using only macroscopic state information.

This limitation, however, can be overcome if we make the following non-dynamical assumption: suppose that the fraction of white or black balls that change colour *at each tick* is equal to the probability μ that an edge has a marker on it, where μ is equal to the number of markers, n, divided by the number of edges, N. That is,

$$\mu = \frac{n}{N} = \frac{w(t)}{W(t)} = \frac{b(t)}{B(t)}. \tag{5}$$

This assumption is the analogue of assuming that the *Stoßzahlansatz* holds *at all times*. We too have posited an absence of correlation. Here, the colour of each ball is taken, at each tick, to be probabilistically independent of whether there is a marker in front of it. We have also introduced a temporally asymmetric element into the model. There are sequences of states of the system that are compatible with the assumption whose time reverse is not. For example, consider a ring that has a white ball at each site and whose markers have been randomly distributed. Now let the system evolve for one tick. All and only the balls that have changed colour have passed a marker. The assumption, which holds for this sequence of states, does not hold for its time reverse. In the later case, ball colours are correlated with the locations of markers. Black balls are found at all and only those sites that have a marker in front of them.

Importantly, the assumption introduced above enables us to express (4) as

$$\Delta(t+1) = \Delta(t) + 2\mu B(t) - 2\mu W(t) = (1 - 2\mu)\Delta(t). \tag{6}$$

This yields,

$$\Delta(t) = (1 - 2\mu)^t \Delta(0). \tag{7}$$

Eqn. (7) is a transport equation, like Boltzmann's equation. It, like Boltzmann's equation, tracks the system's behaviour, including, importantly, its approach to equilibrium. Since $0 < \mu < 1$, (7) tells us to expect $|\Delta(t)| \to 0$ as $t \to \infty$. This is the analogue of Boltzmann's H-theorem. Importantly, this result, like Boltzmann's, is inconsistent with the system's dynamics.

The system's dynamics are symmetric under time-reversal, so not *all* microstates of the system compatible with its macroscopic properties at *any* time lead to a *monotonic* decrease of $|\Delta(t)|$. This is an instance of the reversibility objection. Moreover, because the system is strictly periodic, no initial microstate yields a *monotonic* decrease of $|\Delta(t)|$. This is an instance of the recurrence objection.

Those who intend to more clearly understand the objections, their implications, and their relation to other statistical mechanical ideas, by analysing the model in the way outlined above, are using it to provide an elucidation. If, in proceeding, a user does not intend the model to perform a representational function, then it does not perform a representational function. In such a situation, the Kac ring model is a toy model.

An agent can use the Kac ring to elucidate the recurrence and reversibility objections, without intending that it perform a representational function, because it instantiates a set of properties that are central to the objections. These properties are very general and so are instantiated by many systems, both real and abstract. While it is true that the objections were originally levelled at Boltzmann's dilute gas, there is nothing special about the objections that prohibits an agent from running them against any system that instantiates the set of properties that gave rise to them in the Boltzmannian case, whenever an assumption is introduced that is analogous to assuming that the *Stoßzahlansatz* holds at all times.

What properties are central to the objections? One is that a system possess macroscopic states which are characterised by its microscopic state. Another is that it has a microdynamics that is symmetric under time-reversal. A third is that it displays recurrence. Another is that it permits meaningful talk about equilibrium and nonequilibrium. The objections arise when a suitable assumption is added to systems with these properties. The assumption should posit an absence of correlation at all points in time between certain microscopic parts of the system, and it should enable us to derive, solely on the basis of macroscopic information, a result that predicts a monotonic approach to equilibrium. These properties and this assumption, when taken together, give rise to the objections. They are necessary and jointly sufficient. Since the relevant properties are perfectly general, a number of systems possess them. That is why we can appeal to any system that instantiates all of them to illustrate and elucidate the objections. Naturally, if our purpose is to more clearly understand the objections, then it is reasonable to first choose a simple system, such as the Kac ring, to perform this task. Let us call the set of systems that instantiate all of these properties J. The ring model is a member of J.

All that is needed for an elucidation is that the ring model instantiate the properties that give rise to the objections. Now while it is true that other systems also instantiate these properties, the model's similarity to those systems is not sufficient for it to be said to represent those systems. In particular, just because the Kac ring is similar to each member of J, this does not mean that the Kac ring represents any other member of the set. It also does not mean that the Kac ring represents the set J. The Kac ring is simply a member of J, like Boltzmann's dilute gas.

Using the Kac Ring in a Consistency Proof

In helping agents better understand and appreciate the objections, the Kac ring reveals things about the nature of the introduced assumption and it suggests things about the nature of other *Stoßzahlansatz*-like assumptions. It reveals that the introduced assumption is temporally asymmetric, and it suggests that other *Stoßzahlansatz*-like assumptions are temporally asymmetric. The model suggests that these assumptions cannot hold for all microstates of the systems they apply to at any time and that they cannot hold for all times given any microstate. The Kac ring also suggests that (5) could be true, or approximately true, for times short compared to recurrence times for rings whose initial ball colour distributions are independent of the locations of markers, since there is no mechanism in the system that correlates a ball's colour with markers it has yet to pass. And this idea, in turn, can be thought to suggest situations in which it may be reasonable to make use of temporally asymmetric *Stoßzahlansatz*-like assumptions, and the transport equations they help derive (e.g. Eqn. (7), Boltzmann's transport equation, etc.). In the case of the Kac ring, for rings whose initial ball colour distributions are independent of the distribution of markers, and for times that are short compared to recurrence times. In the Boltzmannian case, for gases comprised of colliding molecules whose initial incoming velocities are uncorrelated, and for times short compared to recurrence times. It is because these and other suggestions so easily emerge from an examination of the Kac ring that make it such a useful hypothesis generating tool.

Why is any of this important? Answer: The reversibility and recurrence objections make it clear that strict interpretations of results such as (7), Boltzmann's equation, and Boltzmann's H-theorem, are untenable. On the other hand, there is something very appealing about them. Eqn. (7), like other transport equations, predicts irreversible behaviour at the macroscopic level. This fits extremely well with experience. Moreover, as David Wallace (2013: p.14) rightly notes, many transport equations, including Boltzmann's, actually work. Boltzmann's equation predicts, quantitatively, and accurately, how dilute gases away from equilibrium actually behave. Naturally, it is worth locating the circumstances in which these equations hold for classical systems that are members of J. This will shed light onto the nature and behaviour of many nonequilibrium systems, both real and abstract. It will help us understand why and how systems with a dynamics that is deterministic, recurrent, and symmetric under time-reversal, exhibit irreversible macroscopic behaviour, and why temporally asymmetric transport equations can accurately predict their behaviour.

Since identifying the circumstances in which transport equations accurately predict the behaviour of actual systems is often very difficult, it can be helpful to use simple models to offer up suggestions of what these circumstances could be. Happily, we can generate reasonable hypotheses about what the circumstances could be for complicated systems on the basis of results that can be shown to hold for simpler systems, such as the Kac ring, say, that are similar to them. While these results do not offer the final word on the matter, they do offer a good way to start the conversation.

The first step in the process is to establish a consistency proof. That is, to show that under certain conditions a system with a dynamics that is deterministic, recurrent, and symmetric under time-reversal, exhibits irreversible macroscopic behaviour—which can be described accurately by a temporally asymmetric transport equation. It is worth noticing that these properties are perfectly general. A number of systems possess them. Both Boltzmann's gas and the Kac ring instantiate these properties. They are similar to one another in these respects. So then, if we discover and can show that under certain conditions the Kac ring consistently instantiates these properties, then we can reason, on the basis of its similarity to Boltzmann's gas, that the gas may also consistently instantiate these properties under these (or similar) conditions by appealing to an argument from analogy (of the form outlined in section 2). Generally speaking, if we discover and can show that under certain conditions the Kac ring consistently instantiates this set of properties, then we can reason, on the basis of its similarity to systems that also instantiate these properties, that these other systems may also consistently instantiate them under these (or similar) conditions by appealing to an argument from analogy. While we are free to consider and test any system with a dynamics that is deterministic, recurrent, and symmetric under time-reversal, that exhibits irreversible macroscopic behaviour, it is, in this context, worth considering the Kac ring first. It is simple, and it has already presented us with conditions under which these properties may be consistently instantiated.

It can be demonstrated that the evolution of an *ensemble* (i.e. a collection of rings that share the same number of sites, N, and initial configuration of black and white balls, but whose number of markers vary) is precisely described by $(1-2\mu)^t$, for very large N, for times much shorter than recurrence times, and whose markers have initially been distributed at random.[22] Importantly, such a result establishes an affirmative answer to the compatibility question. It also suggests truths about the microphysics. Namely that a *Stoßzahlansatz*-like assumption, which is only taken to hold exactly at $t = 0$, in fact contributes to the appearance of irreversible macroscopic behaviour in the model. The result also lends weight to the suggestion of when it is reasonable to make use of temporally asymmetric *Stoßzahlansatz*-like assumptions and the transport equations they help derive. Notice, however, that this result, if valid, is strictly weaker than (7). It claims only that irreversible macroscopic behaviour appears on average. But since the objections show that a strict reading is untenable, this is a happy result.

Consistency Proof: Consider an ensemble in which the number of sites possessed by its constituent rings is very large, say of the order of 10^{23}, ($\sim$the number of molecules that constitute a gas), and consider values of $t \ll N$. Since we are interested in an irreversible move from some nonequilibrium state to equilibrium, suppose, to keep things simple, that the ensemble's initial

[22] Proof of this result appears in a number of texts. See, for example, Kac (1959: Ch.3 Sec.15).

configuration is $B(0) = 0$. To calculate $\Delta(t)$, we introduce $2N$ variables

$$\eta_i(t) = \begin{cases} +1 & \text{if the ball at site } i \text{ at } t \text{ is white} \\ -1 & \text{if the ball at site } i \text{ at } t \text{ is black,} \end{cases} \tag{8}$$

and

$$\varepsilon_i = \begin{cases} +1, & \text{if the edge in front of } i \text{ does not have a marker} \\ -1, & \text{if the edge in front of } i \text{ has a marker.} \end{cases} \tag{9}$$

Then we add up the +1's and -1's for the N sites. Since the colour of the ball at site $i+1$ depends on its colour at i at the previous tick, and whether there was an intervening marker, we have

$$\Delta(t) = \sum_{i=1}^{N} \varepsilon_{i-1}\varepsilon_{i-2}\ldots\varepsilon_{i-t}. \tag{10}$$

If we suppose that markers are set for each ring according to the outcome of a coin toss whose probability of success is $0 < \mu < \frac{1}{2}$, that is

$$Pr\{\varepsilon_j = -1\} = \mu \text{ and } Pr\{\varepsilon_j = 1\} = 1 - \mu, \tag{11}$$

then

$$\langle\Delta(t)\rangle = (1 - 2\mu)^t. \tag{12}$$

This result shows that for very large N and for any $t \ll N$ the evolution of an *ensemble* whose markers are distributed at random and that begins with a white ball at every site is precisely described by $(1-2\mu)^t$. Alternatively, it shows that the evolution of any *ring*, chosen at random from the ensemble, will be approximately described by $(1-2\mu)^t$. Each ring has a dynamics that is symmetric under time-reversal and deterministic. They also display recurrence. Most rings exhibit irreversible behaviour. In fact, irreversible behaviour occurs on average. What we have then, is a model that consistently maintains the properties we are interested in. If, in constructing and establishing this consistency proof, or one like it, a user does not intend for the model to perform a representational function, then it does not perform a representational function. In such a situation, the Kac ring model is a toy model. An agent can use the model in a consistency proof, without intending that it perform a representational function, because it instantiates the set of properties whose compatibility is being tested.

Using the Kac Ring To Generate Hypotheses

The Kac ring has been used to elucidate the reversibility and recurrence objections, to establish the compatibility of certain properties under certain conditions, and to suggest interesting and important things about other systems, including real systems. This section continues to discuss its constructive, hypothesis generating use.

As we have seen, the ring model quite naturally suggests conditions in which other classical systems exhibit irreversible macroscopic behaviour. We were able to generate hypotheses about other systems, using what we learned from the objections, as they applied to the ring model, by reasoning analogically. These hypotheses gained support from the consistency proof offered in the previous section. The success of the proof makes it reasonable to expect a dilute gas to exhibit irreversible macroscopic behaviour for times short compared to recurrence times, if initially the incoming velocities of colliding molecules of a gas are uncorrelated. Moreover, the proof suggests that one would not go too far wrong, in such a situation, if one were to assume that the *Stoßzahlansatz* holds at all times across an appropriate time interval, and were to describe the gas's behaviour using Boltzmann's transport equation. So the Kac ring can be used to help guide the way we think about other systems, including actual systems, such as dilute gases, because it can be used to generate hypotheses about them. It can be used in this way because we can more easily establish results with it and because these results can be linked to systems that instantiate properties shared by the model, by arguing from analogy. Importantly, all of this can be done without intending that the Kac ring represents any or all of these other systems. All that matters is that the ring model instantiates properties that other systems instantiate. These similarities alone permit employing an argument from analogy, whose conclusion we interpret as a hypothesis about the system that shares properties with the ring model. And since similarity is not sufficient for representation and a user's intention is necessary, then if a user does not intend the Kac ring to perform a representational function while generating a hypothesis in this manner, then it does not perform a representational function. In this way, the Kac ring can be used to generate hypotheses about other systems without performing a representational function. And in situations such as these, the Kac ring is a toy model.

Interestingly, it is not only through known similarities (positive analogies) that models, such as the Kac ring, can be used to develop hypotheses about other systems, including actual systems. We can also develop reasonable hypotheses by trading on known dissimilarities (negative analogies).[23]

Here is an example. Kac ring markers are fixed and do not interact with one another. Balls only interact with markers. They do not interact with each other. This means that when markers are uncorrelated with the colours of balls they have yet to encounter, this lack of correlation persists until they come into contact. That is, balls only become correlated with markers once they have passed them. Compare these features to an isolated dilute gas. None of the molecules are fixed and there are no prohibitions on which molecules collide. This means that once the system begins to evolve from an uncorrelated state, correlations almost immediately appear.

[23]See Bartha (2013) for a discussion of positive and negative analogies and their role in analogical reasoning. This terminology was originally introduced in Keynes (1921).

Importantly, correlations emerge between the velocities of molecules that have yet to collide. But we know that a predictively accurate transport equation can be generated for the gas if we assume that the *Stoßzahlansatz* holds at all times (which we know cannot strictly be true) and that a predictively accurate transport equation can be generated for the ring both with and without introducing an analogous assumption. So it is reasonable to think, on the basis of these considerations, that the gas behaves in such a way so as to ensure that the velocities of its molecules behave more like the Kac ring's balls and markers prior to collisions. Since the pre-collision velocities of any pair of molecules could not be uncorrelated, except initially, it appears that they would have to be, what will be called, effectively uncorrelated. That is, that the pre-collision velocities of any pair of molecules would have to be such that it is safe, for the purposes of making accurate predictions about the system's behaviour, to treat them as if they were uncorrelated.

So we have again generated a reasonable hypothesis that concerns actual systems by reasoning analogically. The suggestion that the pre-collision velocities of any pair of molecules of a dilute gas are effectively uncorrelated, at least after some initial time, was arrived at by trading on similarities between the Kac ring and a dilute gas and, importantly, on some of their dissimilarities. This hypothesis generating use of the model depends, in this instance, only on the ring's instantiation of certain properties a dilute gas also instantiates and on the ring's instantiation of properties the gas lacks. These similarities and dissimilarities alone enable us to use an argument from analogy to generate the hypothesis.[24] Since similarity is not sufficient for representation, then if an agent uses the model to generate a hypothesis about dilute gases in this manner and does not intend that it perform a representational function, then the model would not perform a representational function. In such a situation, the ring model would be a toy model.

Some Concluding Remarks

In a recent review article, Frigg and Hartmann (2012: Sec.4.2) have said that toy models:

> . . . are models which do not perform a representational function and which are not expected to instruct us about anything beyond the model itself.

While Frigg and Hartmann are right to claim that toy models do not perform a representational function, it seems misleading to claim that we do not *expect* them to instruct us about anything beyond themselves. What is true is that many of the models that are commonly used as toy models (e.g. the Kac ring, urn model, wind-tree model, etc.) often instruct us about things beyond themselves. As an earlier section highlighted, the Kac ring can be used to elucidate the reversibility and recurrence objections. By examining the Kac ring, one can

[24] In cases like this, however, the following premise is added to the general argument scheme: P3. S is dissimilar to T in certain (known) respects.

develop a more clear understanding of the objections, their implications, and their relation to other statistical mechanical ideas. The ring model can also help us better understand other systems, including actual systems, through its use as a hypothesis generating tool. In testing these hypotheses on actual systems, we can obtain a better understanding of the mechanisms responsible for thermodynamic behaviour and why temporally asymmetric transport equations are predictively accurate.

This paper distinguished toy models from approximations and idealisations. Unlike approximations and idealisations, these models do not perform a representational function. Nonetheless, they play an important role in science. This paper highlighted the important work these models carry out by discussing several ways the Kac ring is used as a toy model. It also explained why the Kac ring can be successfully used in these ways without performing a representational function. Of course, it is natural to wonder about how well this discussion speaks for other toy models. Happily, such a concern only furthers one of the main ambitions of this paper: to encourage a much greater philosophical discussion of these interesting and under-explored models. We should examine how other models are used in ways **1-4**. Only through a much larger discussion will we achieve a richer understanding of toy models and better understand the subtle ways in which they differ from other scientific models. Part of this larger discussion may involve examining and carefully articulating the consequences that follow from the distinction between representational and non-representational models, which thinking about toy models brings out. For example, approximations and idealisations can be, and often are, used to successfully perform function **4**.[25] These models, however, also perform a representational function. Interestingly, this additional function appears to have consequences for the manner in which they are used to generate hypotheses about other systems. Unlike toy models, which, as we saw, generate hypotheses about actual systems via arguments from analogy, approximate and idealised models can be used to generate them directly. If we suppose that the Kac ring represents the generic features of a class of systems (and so is an idealisation), then any conclusion we reach about them by studying the model can be applied directly to any member of the class. Similarly, if we suppose that the Kac ring represents either approximately or in an idealised manner some particular system, then any conclusion we reach by analysing the model can be directly applied to its target.[26]

Hopefully, by way of this entire discussion, this paper has elevated the status of toy models and cultivated a greater interest in them.

[25]In fact, approximations and idealisations can also be, and often are, used to successfully perform functions **1-3**.

[26]Of course, this says nothing about the truth of these inferences. Just that it is reasonable for them to be drawn.

References

Achinstein, Peter. 1968. *Concepts of Science. A Philosophical Analysis.* Baltimore: John Hopkins Press.

Aronson, J. L., R. Harré, and E. C. Way. 1995. *Realism Rescued: How Scientific Progress is Possible.* Open Court.

Bailer-Jones, Daniela M. 2003. "When Scientific Models Represent". *International Studies in the Philosophy of Science* 17: 59–74.

Bartels, Andreas. 2006. "Defending the Structural Concept of Representation". *Theoria* 21: 7–19.

Bartha, Paul. 2013. "Analogy and Analogical Reasoning". In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, fall 2013 edn.

Bokulich, Alisa. 2009. "Explanatory fictions". In *Fictions in Science: Philosophical Essays on Modeling and Idealization*, edited by Maurcio Suárez. London: Routledge, 91–109.

———. 2011. "How Scientific Models Can Explain". *Synthese* 180: pp.33–45.

———. 2012. "Distinguishing Explanatory from Nonexplanatory Fictions". *Philosophy of Science* 79: pp.725–737.

Boltzmann, Ludwig. 1872. "Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen". *Wiener Berichte* 66: pp.275–370.

Bricmont, J. 1995. "Science of Chaos of Chaos in Science?" *Annals of the New York Academy of Sciences* 775: pp.131–175.

———. 2001. "Bayes, Boltzmann and Bohm: Probabilities in Physics". In *Chance in Physics*, edited by J. Bricmont, G. Ghirardi, D. Dürr, F. Petruccione, M-C. Galavotti, and N. Zanghì, vol. 574 of *Lecture Notes in Physics*. Springer Berlin Heidelberg, 3–21.

Brown, Harvey, Wayne Myrvold, and Jos Uffink. 2009. "Boltzmann's H-theorem, its discontents, and the birth of statistical mechanics". *Studies in History and Philosophy of Science Part B* 40: 174–191.

Bunge, Mario. 1973. *Method, Model, and Matter.* Dordrecht: D. Reidel.

Cartwright, Nancy. 1983. *How The Laws Of Physics Lie.* Oxford: Oxford University Press.

Dorfman, Jay Robert. 1999. *An Introduction to Chaos in Nonequilibrium Statistical Mechanics*, vol. 14. Cambridge University Press.

Ehrenfest, Paul, and Tatiana Ehrenfest. 1907. "Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem". *Physikalische Zeitschrift* 8: 311–314.

Elgin, Mehmet, and Elliott Sober. 2002. "Cartwright On Explanation And Idealization". *Erkenntnis* 57: 441–450.

Frigg, Roman, and Stephan Hartmann. 2012. "Models in Science". In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, fall 2012 edn.

Giere, Ronald N. 2004. "How Models Are Used to Represent Reality". *Philosophy of Science* 71: 742–752.

———. 2010. *Explaining Science: A Cognitive Approach*. University of Chicago Press.

Godfrey-Smith, Peter. 2006. "The Strategy of Model-Based Science". *Biology and Philosophy* 21: 725–740.

Goodman, Nelson. 1968. *Languages of Art: An Approach to a Theory of Symbols*. Indianapolis: Bobbs-Merrill.

Gottwald, Georg A, and Marcel Oliver. 2009. "Boltzmann's Dilemma: An Introduction to Statistical Mechanics via the Kac Ring". *SIAM review* 51: 613–635.

Kac, Mark. 1959. *Probability and Related Topics in Physical Sciences*, vol. 1 of *American Mathematical Society*. Interscience Publishers.

Keynes, John Maynard. 1921. *A Treatise on Probability*. London: Macmillan.

Laymon, Ronald. 1990. "Computer Simulations, Idealizations and Approximations". *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1990: pp. 519–534.

McMullin, Ernan. 1985. "Galilean idealization". *Studies in History and Philosophy of Science Part A* 16: 247–273.

Mundy, Brent. 1986. "On the General Theory of Meaningful Representation". *Synthese* 67: 391–437.

Norton, John D. 2012. "Approximation and Idealization: Why the Difference Matters". *Philosophy of Science* 79: pp. 207–232.

Pincock, C. 2012. *Mathematics and Scientific Representation*. Oxford University Press, USA.

Schulman, L. S. 1997. *Time's Arrow and Quantum Measurement*. Cambridge, U.K.: Cambridge University Press.

Suárez, Mauricio. 2003. “Scientific Representation: Against Similarity and Isomorphism”. *International Studies in the Philosophy of Science* 17: pp. 225–244.

———. 2015. “Representation in Science”. In *The Oxford Handbook of Philosophy of Science*, edited by Paul Humphreys. Oxford University Press.

Suppes, P. 2002. *Representation and Invariance of Scientific Structures.* CSLI Publications (Chicago: University of Chicago Press).

Swoyer, Chris. 1991. “Structural Representation and Surrogative Reasoning”. *Synthese* 87: 449–508.

Thompson, Colin J. 1972. *Mathematical Statistical Mechanics.* The Macmillan Company: New York.

Uffink, Jos. 2007. “Compendium of the Foundations of Classical Statistical Physics”. In *Handbook for the Philosophy of Physics*, edited by J. Butterfield and J. Earman. Amsterdam: Elsevier, pp. 924–1074.

Wallace, David. 2013. “What Statistical Mechanics Actually Does”. http://philsci-archive.pitt.edu/9846/.

Weisberg, Michael. 2007. “Three Kinds of Idealization”. *Journal of Philosophy* 104: 639–659.

———. 2013. *Simulation and Similarity: Using Models to Understand the World.* Oxford University Press.

Woodward, J. 2003. *Making Things Happen: A Theory of Causal Explanation.* New York: Oxford University Press.