Check for updates

The Kac Ring or the Art of Making Idealisations

Julie Jebeile[1]

Abstract

In 1959, mathematician Mark Kac introduced a model, called the Kac ring, in order to elucidate the classical solution of Boltzmann to the problem of macroscopic irreversibility. However, the model is far from being a realistic representation of something. How can it be of any help here? In philosophy of science, it is often argued that models can provide explanations of the phenomenon they are said to approximate, in virtue of the truth they contain, and in spite of the idealisations they are made of. On this view, idealisations are not supposed to contribute to any explaining, and should not affect the global representational function of the model. But the Kac ring is a toy model that is only made of idealisations, and is still used trustworthily to understand the treatment of irreversible phenomena in statistical mechanics. In the paper, my aim is to argue that each idealisation ingeniously designed by the mathematician maintains the representational function of the Kac ring with the general properties of macroscopic irreversibility under scrutiny. Such an active role of idealisations in the representing has so far been overlooked and reflects the art of modelling.

Keywords Scientific model · Idealisation · Model explanation · Macroscopic irreversibility · Kac ring · Bolzmann's *H*-theorem · Molecular chaos hypothesis

1 Introduction

Macroscopic phenomena are irreversible. For example, once poured into a same container, two gases do not split again. A coin dropped on the floor remains there, we see no coin jumping up by itself from the floor. And yet macroscopic phenomena result from a microscopic dynamics described by reversible laws, i.e., laws invariant under time transformation.

✉ Julie Jebeile
  julie.jebeile@gmail.com
  https://www.juliejebeile.net/en

[1] Institute of Philosophy & Oeschger Center for Climate Change Research, University of Bern, Bern, Switzerland

Ludwig Boltzmann offered in 1872 an explanation of why macroscopic phenomena are irreversible despite having a reversible microscopic dynamics: this is the *H*-theorem. However, the *H*-theorem has been plagued by two important paradoxes that lie at the foundations of statistical mechanics: the "reversibility paradox" and the "recurrence paradox".

That is the reason why, later on, in 1959, mathematician Mark Kac introduced a model, called the Kac ring, as an "analog" of the classical solution of Boltzmann to explain macroscopic irreversibility ([27], p. 99). The Kac ring aims "to reconcile both time reversibility and recurrence with "observable" irreversible behavior" ([27], p. 73), and to offer and evaluate a proper interpretation of Boltzmann's molecular chaos hypothesis. Thereafter it has undergone multiple developments (e.g. [10]), especially in quantum mechanics (e.g. [11–13, 41]).

But the Kac ring is far from being a realistic representation of something. It is a simple ring with N sites; some of them are said to be active. On each site, there is a black or a white ball. Every ball moves counterclockwise to the next site. When a ball leaves an active site, it switches colour. How can such a model be of any help here?

Philosophical accounts of how models, albeit idealised, can teach us something about the world were first developed in the 1990s (e.g., [30, 31, 34]) and, since then, have flourished (e.g., [3, 4, 22, 28, 38, 42]). Most of these accounts argue that models can provide explanations of the phenomenon they are said to approximate, in virtue of the true components they contain, and in spite of the idealisations they are made of. On this view, idealisations are not supposed to contribute to any explaining, and should not affect the representational function of the other model components.

As I show in this paper,[1] these accounts fail to give reasons why the Kac ring is actually used trustworthily to understand the treatment of irreversible phenomena in statistical mechanics while it is only made of idealisations (Sect. 4). My aim is to show that the reasons bear on the very idealisations: each idealisation ingeniously designed by the mathematician maintains the representational function of the Kac ring with the general properties of macroscopic irreversibility under scrutiny (Sect. 5). Yet such an active role of idealisations in the representing has so far been overlooked by philosophers of science interested in models and idealisations. By investigating the Kac ring, I endeavour to reveal Mark Kac's art of designing adequate idealisations that enable the users to draw information *easily* and *trustworthily*.

Beforehand, I first present the historico-scientific context of the creation of the Kac ring (Sect. 2), and make it clear how the model reconciles both time reversibility and recurrence with "observable" irreversible behaviour, and enables us to evaluate interpretations of Boltzmann's molecular chaos hypothesis (Sect. 3).

---

[1] This work is drawn from Jebeile [24], Part I, Chap. 3.

2 Historical Context

2.1 The *H*-Theorem

In order to explain macroscopic irreversibility, Boltzmann formulated the *H*-theorem in 1872. He considered a non-equilibrium ideal gas and assumed that the gas particles interact mechanistically by means of repulsive short-range and attractive forces.

He introduced a new physical magnitude, the function H that is related to the entropy S by the equation $H = \frac{-S}{N.k_B}$ for a gas with N particles; k_B being the Boltzmann constant. The *H*-theorem states that H decreases monotonically over time and then remains constant when the particle density in the system reaches the value at equilibrium.

Boltzmann crucially added the molecular chaos hypothesis (or *Stosszahlansatz*): it means that there is no correlation between the velocities of two particles that are bound to collide against each other. This hypothesis is notably based on that the particle distribution is supposed to be spatially homogeneous ([16], 3.b, 6). It follows that the number of collisions between two groups of particles during the period of time δt is proportional to their respective particle density.

That way, the *H*-theorem singles out one time direction. And the molecular chaos hypothesis is supposed to play a key-role to explain irreversibility.

### 2.2 The Reversibility Paradox and the Recurrence Paradox

The *H*-theorem has been plagued by two important paradoxes that lie at the foundations of statistical mechanics: the "reversibility paradox" and the "recurrence paradox".

The *reversibility paradox* (*Umkehreinwand*) was expressed in 1876 by Johann Loschmidt. If a gas starts from the initial state s_0 and reaches the state s_t after time t, then, according to the *H*-theorem, $H_t \leq H_0$. Now if the velocity of every gas particle is reversed, the gas would come back to its initial state $s_{t'} = s_0$ after time $t'(= t)$, and H would still continue to decrease so that $H_{t'} = H_0 \leq H_t$. Therefore H cannot decrease over time but remains constant, which contradicts the *H*-theorem.

The *recurrence paradox* (*Wiederkehreinwand)* was expressed in 1896 by Ernst Zermelo. It is based on Poincaré's recurrence theorem according to which a closed dynamical system comes back arbitrarily near its initial state after a sufficiently long period of time. H thus cannot always decrease over time and is expected to be periodic.

According to Kac [27], Boltzmann replied to Loschmidt's objection: "go ahead, reverse them!". Against Zermelo's objection, he argued that the time periods, i.e. the Poincaré cycles, are too long for the recurrences to be observed. But, in spite of Boltzmann's responses to the paradoxes ([6–8]), it was not until the Paul and Tatiana Ehrenfest's model ([15]) and the Kac ring model that one could see the actual

Fig. 1 Ring model with $N = 25$
and $A = 9$
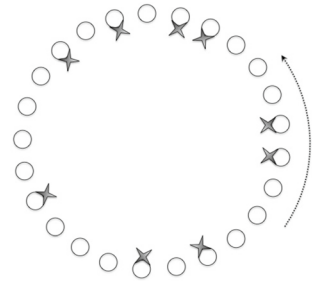
possibility to reconcile both time reversibility and recurrence with "observable" irreversible behaviour.

2.3 The Kac Ring

The Kac ring model aims at deriving an irreversible macroscopic behaviour from a reversible microscopic dynamics and at illustrating the two paradoxes.

As said, the model is a ring with N sites. On each site there is a black or a white ball. Every ball moves counterclockwise to the next site. There are some fixed active sites, called "scatterers". A is their number. When a ball leaves an active site, it switches colour, turning white if it was black, black if it was white. Figure 1 illustrates a Kac ring model with $N = 25$ and $A = 9$; the active sites are marked, colour of balls is not represented.

The macroscopic description of the Kac ring is given by the *overall colour* of the ring. The ring is supposed to be at equilibrium when the overall colour is "grey": it is composed of half black balls and half white balls.

$W(t)$ is the number of white balls at time t and $B(t)$ the number of black balls. $W_A(t)$ is the number of white balls on the active sites at time t and $B_A(t)$ the number of black balls on the active sites.

The numbers of white and black balls are respectively given by:

$$W(t + 1) = W(t) - W_A(t) + B_A(t)$$
$$B(t + 1) = B(t) - B_A(t) + W_A(t) \tag{1}$$

The overall colour of the ring δ equals the normalised discrepancy:

$$\delta(t) = \frac{W(t) - B(t)}{N} \tag{2}$$

$\delta = 1$ means that the ring is "white" (i.e. only composed of white balls); $\delta = -1$ means that the ring is "black" (i.e. only composed of black balls); and $\delta = 0$ means that the ring is "grey" (i.e. half black, half white) and therefore at equilibrium.

The translation of Boltzmann's molecular chaos hypothesis into the Kac ring implies that the colour of each ball is not correlated to the fact that the site on which the ball stands is active and is going to make the colour of the ball change. Therefore

the number of white (or black) balls on active sites equals the total number of white (or black) balls multiplied by the fraction of active sites $\mu = \frac{A}{N}$:

$$
\begin{aligned}
\frac{W_A(t)}{W(t)} &= \mu \\
\frac{B_A(t)}{B(t)} &= \mu
\end{aligned}
\tag{3}
$$

3 Uses of the Kac Ring

3.1 Explaining Observable Irreversible Behaviour

In a series of lectures published under the title "*Probability and Related Topics in the Physical Sciences*" ([27]), Kac aims to explicate the interrelation of dynamical law, probability, and initial conditions in physics, and notably to introduce statistical mechanics of irreversible phenomena. In this context, he endeavours to elucidate "Boltzmann's not always clear ideas" ([27], p. 72).

After discussing the virtues of the P. and T. Ehrenfest model, "probably one of the most instructive models in the whole of Physics" ([27], p. 73), Kac concludes that "In spite of the brilliant analysis of P. and T. Ehrenfest which elucidated Boltzmann's not always clear ideas and made them highly plausible, the question [of macroscopic irreversibility] even now (45 years after the appearance of Ehrenfest's work!) is not fully settled" ([27], p. 72).[2] As Kac highlights, "There are broadly (and somewhat vaguely) speaking two problems:"

> I. Is it possible to reconcile both time reversibility and recurrence with "observable" irreversible behavior?
> II. Is it possible to achieve such a reconciliation in the realm of classical mechanics? ([27], p. 73)

For Kac, the first problem can receive an affirmative answer "if a model is found which exhibits the required features", and specifically "P. and T. Ehrenfest proposed such a model in 1907" ([27], p. 80). The second problem is more fundamental, and the Kac ring is aimed "to understand [this] more basic problem of how probability enters classical mechanics" ([27], p. 99). Hilariously, Kac justifies the adequacy of the ring model for this purpose by claiming that the ring model is "another artificial model which in spirit is much closer to "reality""([27], p. 99).

Given that "The objections of Loschmidt and Zermelo made it clear that the naive formation of the H-theorem is untenable", Kac wants to demonstrate that "The "Stosszahlansatz" [...] on which the derivation of Boltzmann's equation rests cannot be a purely dynamical conclusion", and that "hence some reinterpretation

---

[2] In the P. and T. Ehrenfest model, "2R balls numbered from 1 to 2R are distributed in two boxes A and B. An integer from 1 to 2R is chosen "at random" and the ball corresponding to that number is moved from the bow in which it is to the other one. The process is then repeated a desired number of times"([27], p. 73).

is necessary", "Boltzmann himself [having] proposed that the *H*-theorem be inter-preted statistically" ([27], p. 72). Kac therefore introduced the ring model in order to evaluate a Boltzmann interpretation and a Gibbs interpretation of the molecular chaos hypothesis, and finally to promote the latter.

> "The Boltzmann entropy behaves in a much more erratic way [...] and only *on the average* can it exhibit monotonic behavior. [...]
> All this points to two quite different formulations of the second law of thermo-dynamics.
> In one formulation (Boltzmann), the notions of state and entropy are quite intuitive but the monotonic increase of entropy cannot be strictly maintained.
> In the other (Gibbs) the increase of entropy is a rigorous theorem but the notions of entropy and approach to equilibrium are much less intuitive and direct."([27], pp. 98–99, *emphasis added*)

Interestingly, the ring model is used to consider each of the two interpretations. First, it is used as the "analog of the Classical Solution of Boltzmann." Therefore, the analog of the molecular chaos hypothesis, given by Eq. (3), is introduced in the system of Eqs. (1) and (2), and leads to:

$$\delta(t+1) = \frac{W(t+1) - B(t+1)}{N} = (1 - 2\mu)\delta(t) \tag{4}$$

By recursion it follows that:

$$\delta(t) = (1 - 2\mu)^t \delta(0) \tag{5}$$

Thus, under the analog of the molecular chaos hypothesis, the overall colour of the ring is given by (5) and follows a monotonic behaviour. μ takes a value between 0 and 1, consequently $\delta(t)$ is a decreasing function (alternating in sign if $\mu > 1/2$). $\delta(t)$ has a limit equal to zero when t tends to infinity; at zero, the ring is "grey" and therefore at equilibrium.

The Kac ring thus exemplifies that the molecular chaos hypothesis entails a monotonic behaviour of a macroscopic system having a reversible microscopic dynamics, but illustrates at the same time that this hypothesis can be objected by the reversibility paradox and the recurrence paradox.

> "...we obtain a monotonic approach to equipartition of white and black balls.
> This conclusion is clearly untenable because our model is completely "reversible."In fact starting with all white balls [...] we continue for a while and then reverse the colors of the balls; let them move clockwise and move all elements of [the sites] one unit counterclockwise. We shall arrive then at the initial state contrary to [5] which ought to hold for the "reversed"model as well as for the original one.
> Moreover, the model is strictly periodic with period [2*N*] (Poincaré cycle) which is again incompatible with [5]." ([27], p. 100)

With the help of the ring model, Kac further promotes the probabilistic interpreta-tion, i.e. the Gibbs interpretation, instead. For him, Eq. (5) can be justified "on the

average" ([27], p. 102). "It is perfectly clear that the origin of this difficulty is in formulas [3, i.e. the Stosszahlansatz" formula]. These formulas must be interpreted and their meaning analyzed more carefully" ([27], p. 102).

For his demonstration, Kac introduces another notation. The site variable ϵ_p is governed by the following rules:

$\epsilon_p = -1$ if $p \in A$

$\epsilon_p = +1$ if $p \notin A$ where A is the set of active sites.

The state of the ball p is determined by the dynamic variable η_p:

$\eta_p(t) = +1$ if the ball on the site p at time t is white

$\eta_p(t) = -1$ if the ball on the site p at time t is black

During a normal run, the balls move counterclockwise, and thus the time evolution of the ring is given by $\eta_p(t) = \eta_{p-1}(t-1)\epsilon_{p-1}$.

By recursion, we obtain: $\eta_p(t) = \eta_{p-t}(0)\epsilon_{p-1}...\epsilon_{p-t}$.

Thus $\delta(t) = \frac{1}{N}\sum_p \eta_p(t) = \frac{1}{N}\sum_p \eta_{p-t}(0)\epsilon_{p-1}...\epsilon_{p-t}$.

Originally, Kac then considers the case $t = 0$ with $B(0) = 0$ and therefore $\eta_p(0) = +1$, but let us follow Schulman [37]'s analysis and continue with the general case. The expression of $\delta(t)$ therefore becomes:

$$\delta(t) = \frac{1}{N}\sum_p \eta_p(t) = \frac{1}{N}\sum_p \eta_p(0)\left[\prod_{l=0}^{t-1}\epsilon_{p+l}\right] \tag{6}$$

If we assume that all possible positions or the set of sites are equiprobable, then we can calculate the average of the previous formula. In this case, the average of the dynamic variables on the products of site variables—which correspond to the sum of the terms over p in Eq. (6)—can be approximated by using the average of the site variables:

$$\sum_p \left[\prod_{l=0}^{t-1}\epsilon_{p+l}\right]\eta_p(0) \Rightarrow N \times \left\langle\left[\prod_{l=0}^{t-1}\epsilon_{p+l}\right]\right\rangle\eta_p(0) = N \times \left[\prod_{l=0}^{t-1}\langle\epsilon_{p+l}\rangle\right]\eta_p(0) \tag{7}$$

And the calculation of $< \epsilon >$ becomes:

$$< \epsilon > = (+1)Pr(\epsilon = +1) + (-1)Pr(\epsilon = -1)$$
$$< \epsilon > = (+1) * \frac{N-A}{N} + (-1) * \frac{A}{N} \tag{8}$$
$$< \epsilon > = 1 - 2\mu$$

Introducing (7) and (8) in (6) leads back to Eq. (5), i.e. $\delta(t) = (1-2\mu)^t\delta(0)$. In this derivation, the assumption of equiprobability plays an equivalent role to the *Stosszahlansatz*. It specifies a probability distribution, and is necessary for a probabilistic treatment of the problem. "As convincing as this argument is, it should be remembered that it contains an element of arbitrariness in the assumption that all positions of the set [of sites] are equally probable. Although undoubtedly this assumption can be weakened, the fact remains that some such assumption must be made"([27], p. 102). In other terms, the "*Stosszahlansatz*" can be justified "on the average".

Furthermore, the purpose of Kac (but also Boltzmann, Gibbs and Maxwell) is to show that there is no sharp dichotomy between reversibility and irreversibility but rather an increasing improbability of some trajectories of the system. The improbability is higher as the system's degrees of freedom increase in comparison to the observation time, up to the point of becoming a practical impossibility (see, e.g. [33], pp. 279–280). The Kac ring helps us seeing this gradual transition through the parameter N. The important conclusion is that, for large N, large t, but small $\frac{t}{N}$, it is amazingly improbable to observe a system with a *non-monotonic* approach to equipartition (i.e. $\delta(t) = 0$); thus every *individual* system will show a monotonic approach.[3]

> Suppose we plot $[\delta(t)]$ against t for each set [of sites]. We obtain $[C_A^N]$ "curves"; all start from 1 at $t = 0$ and all have period $[2N]$.
> Suppose we are going to look on these curves at a fixed point $t \ll [N]$. Think of $[N]$ being of order 10^{23} and t of order 10^6. At t the ordinates of all curves concentrate very strongly about $[(1 - 2\mu)^t]$—it takes extremely bad luck to observe a sizable deviation from $[(1 - 2\mu)^t]$! ([27], p. 102)

In a nutshell, the ring model enables the evaluation of two interpretations of the problem of macroscopic irreversibility. On one side, the model shows the limits of the Boltzmann interpretation. The analog of Boltzmann's molecular chaos hypothesis does not hold for any actual set of active sites, even if the distribution of active sites is set randomly. For instance some microscopic configurations do not reach equilibrium. Bricmont [9] takes the example of a ring that is initially composed of only white balls, and has alternatively active and non-active site. After two time steps, all the balls become black. Two time steps later, they all become white, and so one. In this case, the ring is four-periodic and never reaches equilibrium. But the molecular chaos hypothesis is expected to hold "on average" in that it represents to some extent the typical behaviour of large sized rings. The monotonic behaviour of $\delta(t)$ is also valid for very large rings insofar as the equation is a good approximation for a long period of time. On the other side, the model enables the appreciation of the relevance of the Gibbs interpretation. It can be used to describe no ring in particular, but rather an averaged set of rings that all share the same macroscopic features, i.e., N, μ and $\delta(0)$, but have sites arranged differently. The average behaviour is here predominantly irreversible.

3.2 Illustration of Time Reversibility and Recurrence

I now want to make it clear how the model illustrates the reversibility paradox. During a normal run, the ring becomes more and more grey. Let us suppose that we reverse the dynamics. Then the rules change: the balls move clockwise to the next site and switch colour when they *arrive at* an active site.

---

[3] I am in debt with one of the two reviewers who made this point very clear to me and provided the reference.

Since the expression (5) of $\delta(t)$ remains the same, it is expected to decrease and the ring to become more and more grey. But, after several time steps T, the reversed process leads the ring back to the initial state. Indeed the ring goes through the same states as during the normal run; the series of states the system goes through is symmetrically identical to the series of states it went though the first time.

Here the model not only exemplifies time reversibility but also illuminates the action of the molecular chaos hypothesis: In the reversed run, there is a correlation, for a ball, between "having a colour"and "having an active site behind itself". Each ball "revisits" in the opposite direction the sites of the ring on which it went through already. In other words, the future of the ring is absolutely correlated to its past. As a result, Eq. (5) becomes invalid and the reversibility paradox irrelevant.

Using the second notation, the reversed run can be given by $\eta_p(t) = \epsilon_p \eta_{p+1}(t-1)$. At time $t+1$, when the rotation of balls is modified, we thus obtain $\eta_p(t+1) = \epsilon_p \eta_{p+1}(t) = \epsilon_p \epsilon_p \eta_p(t-1)$ where the variable ϵ_p appears twice (note that its square always equals one). In other words, the site variables are correlated.

It is important to note that the ring model supports a subtle definition of time reversibility. Usually, time reversal is supposed to follow the operation "$t \to -t$". But time reversibility for dynamical systems is better expressed as follows: the dynamics of a system with a state space S is time reversible if and only if: if $t \to S = Q(t)$ is a solution to the equations of motion, then $t \to S = \pi[Q(-t)]$ is also a solution to the equations of motion. π is a parity transformation that depends on the specific dynamical system under scrutiny. In point-mass mechanics, for example, with generic state (x_i, p_i), it is defined as $\pi[(x_i, p_i)] := (x_i, -p_i)$. Thus, in the Kac ring, it is the ball's colour that actually reflects the momentum's parity transformation. As Kac writes "our model is completely "reversible." In fact starting with all white balls [...] we continue for a while *and then reverse the colors of the balls*; let them move clockwise and move all elements of [the sites] one unit counterclockwise" ([27], p. 100, *emphasis added*).[4]

The model also illustrates the recurrence paradox since it is $2N$-periodic. Let us suppose that the ring is initially only composed of white balls and evolves during $2N$ time steps. Every ball went twice on the active sites. The system has therefore switched colour two consecutive times. Hence it has become fully white again. Whatever the initial situation is, after $2N$ time steps, the system thus finds its initial state back.

The model also offers a possible resolution to the recurrence paradox, which exemplifies Boltzman's own reply: If the number of sites N on the ring is extremely high, one could not observe a reversible behaviour. The reason is that the time for recurrence to occur is proportional to N and, if N is extremely high, the time would be too long for the user to observe recurrence.

---

[4] I am also in debt with one of the anonymous reviewer who points to this aspect.

3.3 Trustworthiness of the Model

In a nutshell, the ring, characterised by the overall colour $\delta(t)$, enables us to grasp how both time reversibility and recurrence can be reconciled with "observable" irreversible behavior, and to evaluate the relevant interpretation of the molecular chaos hypothesis in the explanation of the irreversibility of the macroscopic behaviour of the ring. But can we take the conclusions drawn from the model for granted?

Among scientists, there is much disagreement about whether the Kac ring explains macroscopic irreversibility (see [14], p. 58). But this is *not* because the Kac ring is a misrepresentation. The model is considered as trustworthy for the intended use, e.g., to exemplify and to evaluate Boltzmann's solution.

Thus the model has further been developed to test the molecular chaos assumption. Among Boltzmann's hypothesis, it is the only assumption to be probabilistic (the others are deterministic), and therefore to be non-mechanical. Coppersmith has tried to find another *origin* of irreversibility with the Kac ring. He inserted non-time-reversal invariant scatterers in the model for proving that these "anomalous" scatterers are responsible for irreversibility ([10]). Fernando claims that the "true origin" of irreversibility lies on the stochastic interaction between the system and the environment whereas the molecular chaos hypothesis "for deterministic dynamics helps to conceal rather than reveal the true origin of irreversibility" ([18], p. iii). Other physicists have used the model to study the probabilistic nature of the microscopic dynamics (see [13, 21]).

The reasons of disagreement expressed in the relevant scientific literature blame instead the underlying theoretical hypotheses. Importantly, the molecular chaos hypothesis is controversial ([1]). Some scientists consider that it provides the temporal asymmetry of thermodynamical behaviours in an *ad hoc* way rather than *explaining* macroscopic irreversibility.

We also need to understand why the initial states of the ring are of low entropy since external human operation is needed to prepare such unusual initial states of isolated systems. Thus, for Schulman ([37], p. 29), the model does not succeed in deriving the arrow of time. Others, e.g. Bricmont [9] and Price [36], have also insisted on that initial conditions may contribute in explaining irreversibility.

Another source of disagreement concerns the way the reversed dynamics of the ring is induced ([14]). Doubts have indeed arisen about whether this operation is purely mental or involves an external action on the system. And yet if it is an external action, there would be a negative entropy from the environment that would constitute an important component to explain the second law of thermodynamics.

Since the Kac ring is used trustworthily to understand certain aspects of statistical mechanics of irreversible phenomena, it is worth investigating the very reasons why we are warranted in trusting the model for this purpose. For that matter, let us turn to the current philosophical accounts of why scientific models, albeit idealised, can provide explanations of the phenonemon they are said to approximate. As we will see, most of these accounts assume that model idealisations do not contribute to any explaining. After a brief review, my aim will be to show that, contrary to what these accounts tell, the success of the Kac ring stands in its very idealisations.

4 Characterisation of the Ring: What Makes it Trustworthy?

4.1 Targetless Toy Model

The Kac ring represents general properties of macroscopic irreversibility. Nonetheless it can be said to be targetless in that it does not aim to describe accurately a real system, process or phenomenon in particular. Targetless models are some kind of abstract direct representations ([43]). In targetless models, such as cellular automata like the Game of Life or crystal growth, "no target is chosen at all. The only object of study is the model itself, without regard to what it tells us about any specific real-world system. This type of modelling is most akin to pure mathematical analysis." ([44], p. 129).

For Luczak [32], because the Kac ring is targetless, it follows that it does not perform a representational function. Luczak argues that the Kac ring is a toy model and, "Unlike idealisations and approximations, toy models do not perform a representational function. That is, they do not represent anything." (*Ibid.*, 1).[5] This nevertheless seems to stand in contradiction with the fact that the Kac ring is a good pedagogical device or a heuristic tool that, like any other toy model, enables us "To learn to use, or to become comfortable with, certain formal techniques", "To elucidate certain ideas relevant to a theory", "To test the compatibility of various concepts", or "To generate hypotheses about other systems" ([32], p. 1).

This may be a merely terminological dispute though, since Luczak concedes that toy models can be used by agents because they "instantiate" properties; instantiation being obtained by *similarity* with systems. At some point, he also writes that "If we suppose that the Kac ring *represents* the generic features of a class of systems [...], then any conclusion we reach about them by studying the model can be applied directly to any member of the class." (*Ibid.*, 7, *emphasis added*). But Luczak's view does not tell *what* does the "instantiating" in the model that is only made of idealisations, and *why* we can be confident in the "instantiating" and thereby *why* we can trust the Kac ring for the purpose of understanding the statistical mechanical treatment of irreversible phenomena.

My work consists precisely in examining how each idealisation maintains the representational function of the Kac ring with the general properties of the class of irreversible macroscopic phenomena. Beforehand, I need to show that most philosophical accounts of how models can explain fall short in describing how the Kac ring works. These accounts consider models either as approximations, minimal models, or caricatures.

---

[5] Following Giere [20], Luczak adopts the perspective of the intentional agents that use models for what counts as a representation. For Luczak, "it seems fair to say that a model performs a representational function only if its user intends for it to perform a representational function [...] it is not the model that is doing the representing; it is the scientist using the model who is doing the representing." ([32], p. 2).

4.2 Approximation

Many accounts have in common the thesis according to which scientific models can teach us things about actual empirical systems if they are approximate descriptions, or *approximations*, of their target. Here models must represent their target systems to some degree of accuracy and therefore distort as little as it is compatible with tractability.

For most of these accounts, a model represents the target phenomenon:

(i)   in spite of having idealisations that should be employed without jeopardising the representational function of the model, and

(ii) in virtue of having certain components that mirror/denote/are isomorphic to/ are similar to/partially resemble/approximate/are true about/etc. the relevant features of the target system.

Generally each author commits herself to an account of the representational relationship and its success (whether they rest on isomorphism, similarity or denotation, for instance).

The de-idealisation thesis is one of these accounts. It is most famously represented in McMullin [34], but is also present in Laymon ([30, 31]) and Nowak ([35]). It describes a de-idealisation process that is about adding features of the target back in to models and correcting mathematical approximations that appear in the original models. According to the de-idealisation thesis, models perform a representational function if they are approximately true. We know that they are approximately true if, once de-idealised, they provide more accurate predictions. In this thesis, it is assumed that their "true" part captures what Mcmullin ([34]) calls the "real structure" of the target system, and it necessarily excludes the idealisations.

According to a more recent account (Bokulich's [3, 4]), models are explanatory if they correctly capture the structure of counterfactual dependences in the target system, i.e. if they express the physical relationships between variables of interest adequately so to answer a broad class of "what-if-things-had-been-different" questions under some changes of initial conditions. For that, they must be in an isomorphic relation with the relevant features of their targets.

For some others, models are good representations if, apart from idealisations, they preserve a minimal amount of relevant accurate information about their target. The idealisations can help in isolating the causal factors of interest within the model. For instance, Strevens ([38]) argues that idealisations are used to highlight within a model those factors that are not causally relevant to the explanation of the target.

Bokulich's and Streven's are not traditionally taken to be accounts of models as approximations. However, they both endorse the view that, within the model, true elements are doing the representing while idealisations play another role.

There is no point in analysing the Kac ring as an approximation. Such accounts fall short in explaining how the Kac ring works for the following obvious reasons:

First, it does not make sense to consider the possibility of de-idealising the ring. The ring meets its purpose precisely in virtue of its idealisations. De-idealising the

ring would lead back to the (already idealised) representation of the gas system underlying the *H*-theorem; therefore there would be no progress made here.

Second, there is no "true" component within the Kac ring. The model does not contain accurate elements about the target that would perform solely the representational function. The Kac ring is helpful precisely because it idealises, rather than abstracts away, features from the target.

Third, the above-mentioned accounts describe how models explain in spite of having idealisations. But all the components in the Kac ring are idealisations.

4.3 Minimal Model

We may hypothetise that the Kac ring is a minimal model as defined by Batterman and Rice ([2]). Minimal models "are used to explain patterns of macroscopic behavior across systems that are heterogeneous at smaller scales" ([2], p. 349); "these minimal models are explanatory because there is a detailed story about why the myriad details that distinguish a class of systems are irrelevant to their large-scale behavior. This story demonstrates, rather than assumes, a kind of stability or robustness of the large-scale behavior we want to explain under drastic changes in the various details of the system." ([2], p. 373)

The Kac ring, like minimal models, is not explanatory in virtue of the features it has in common with real systems, since it has no such feature in common. It may also be considered as a way of delimiting a universality class, that would be here a set of general properties underlying macroscopic irreversibility.

Unlike minimal models, though, the initial purpose of the model is not to demonstrate that the details that distinguish the model and various real systems are irrelevant, even though this may be seen as a by-product of the model.

More importantly, the Kac ring is not devoid of representation relations. Denying the Kac ring a representational function would not do justice to the work and creativity of Mark Kac in designing the appropriate idealisations of the ring.

4.4 Scientific Caricature

Gottwald and Oliver suggest that the Kac ring model is a caricature when they claim that the model provides "a caricature of the underlying issues while being simple enough that they can be exposed through elementary and explicit computation" ([21], p. 614). Let us take this idea seriously here.

In an influential paper ([19]), Gibbard and Varian distinguish scientific caricatures from models as approximations. Although Gibbard and Varian focus on economic modelling, their definitions can be made general. While approximations aim to describe reality, albeit in an approximate way", caricatures "seek to "give an impression" of some aspect of economic reality not by describing it directly, but rather by isolating and exaggerating certain selected aspects of the economic situation" ([19], p. 665).

While idealisations in approximations are due to simplicity and tractability, scientific caricatures emphasise, even distort, some features of the target system and

omit others for the purpose of isolating and illuminating the effects of the factors involved.

> Although, [...] when a model is applied as a caricature, it may indeed be hypothesized that the model is an approximation of the roughest kind, the model will be chosen not for the sake of good approximation, but to distort reality in a way that illuminated certain aspects of that reality. ([19], p. 676)

This is similar to a pictorial caricature. It emphasises, even to the point of distorting, some features of the subject and omit others for the purpose of giving the reader (or the viewer) immediate adjacent information about the subject. For example, the size of a subject's nose may be exaggerated in a pictorial caricature, so to bring the attention of the audience to certain aspects of the subject's character ([17], p. 10).

But the Kac ring is worse than a caricature as it contains no component aiming at representing realistically a specific empirical target in particular. That said, the idealisations involved in the Kac ring are as active as the ones in caricatures. They capture important aspects of the idealised gas representation underlying the *H*-theorem. These components are:

(i)   the relation between the macroscopic and microscopic levels of the system,
(ii) the multi-realisability of the macroscopic behaviour,
(iii) the molecular chaos hypothesis,
(iv) the reversibility and the recurrence of the microscopic dynamics.

They are the most relevant aspects for the purpose of illustrating macroscopic irreversibility, deriving the reversibility and recurrence paradoxes, and considering how they relate to Boltzmann's *H*-theorem.

4.5 Analog

The Kac ring is not an approximation, not a caricature, nor a minimal model. The idealisations it contains play an active role similar to the one defined within the account of models as caricatures.

How to characterise the relationship between the Kac ring and the targeted properties of macroscopic irreversibility? We should take seriously the claim that the model is, as Mark Kac initially introduced it, an "analog of the classical solution of Boltzmann"([27], p. 99).

Here the term "analog"does not designate what is traditionally meant by "analogy" in philosophy of science, i.e., a mathematical representation that connects different phenomena. A famous example is the analogy between waves of light, sound and water ([23], pp. 10–12). The equation $y = a \sin 2\pi f x$ can be used in three cases. It can describe the height of a water wave at point x; in that case, a is the amplitude and f *is* the frequency of the waves. It can describe the amplitude of a sound wave at point x; here, a is the loudness and f is the pitch. Or it can describe the amplitude of a light wave at point x; in that case, a is the brightness and f is the colour.

The reason why this is not an analogy is that the mathematical equation in the Kac ring is not strictly the same as the Boltzmann transport equation. That said, it mathematically imitates the behaviour described by Boltzmann.

On one side, Boltzmann distinguishes two gases: one is composed of particles A whose velocities are $\mathbf{v}$ inside the differential element $d\mathbf{v}$, and the other is composed of particles B whose velocities are $\mathbf{w}$ inside $d\mathbf{w}$ ([27]). A collision occurs when A (or B) meets B (or A). After a collision, $\mathbf{v}$ and $\mathbf{w}$ respectively become $\mathbf{v} + (\mathbf{w} - \mathbf{v}).\mathbf{ll}$ and $\mathbf{w} - (\mathbf{w} - \mathbf{v}).\mathbf{ll}$. These changes are calculated from the momentum conservation and from the energy conservation for elastic collision.

On the other side, in the Kac model, the balls can be either black or white. They switch colour when they meet (or leave) an active site. One thus obtains $W(t + 1) = W(t) - W_A(t) + B_A(t)$ and $B(t + 1) = B(t) - B_A(t) + W_A(t)$. The analog of the Boltzmann transport equation is $\delta(t + 1) = \frac{W(t+1)-B(t+1)}{N} = (1 - 2\mu)\delta(t)$.

In the Kac ring, each idealisation plays an active role in the representing by being an analog of a general property of macroscopic phenomena as hypothetised in the H-theorem of Boltzmann. I now want to reconstruct *how* the model idealisation does so. My intention, though, is not to reconstruct the genesis of how the model has been historically derived. The Kac model was clearly influenced by the Ehrenfests' earlier models (i.e., urn but also wind-tree), and could also be seen as a further adaptation of these models ([5] discusses similar cases). But my account is synchronic: I examine each idealisation in the model and explains how it plays a role in both the intelligibility and the acceptability of the model.

5 Active Role of Idealisations

The idealisations in the ring are designed to facilitate the inferential work, i.e. they make consequences of the Boltzmann's assumptions easy to draw. Exemplifying Boltzmann's solution would have been more difficult if one had to take into account the actual complicated distribution of the particles in a gas. The inferential capacity of models has received much philosophical attention these last decades (e.g., [20, 25, 29, 39, 40]).

The success of the idealisations stands not only in that they facilitate the inferential work, but also in that, despite being distortions, they capture important general properties of the (already idealised) gas representation underlying the H-theorem.

5.1 Descriptions of Macroscopic and Microscopic Levels

A first important feature that the Kac ring captures is that the macroscopic behaviour of the gas is derived from its microdynamics. In the idealised representation underlying the H-theorem, the gas is composed of a high number of particles, represented by *hard spheres of a certain diameter* that can collide against each other. The macroscopic behaviour of the gas derives from the individual trajectories of the particles. It is characterised by the function H whose calculation depends on the velocities of each particle.

In the Kac model, the derivation of the macroscopic behaviour from the microdynamics is captured in the following way. The ring is composed of microscopic entities, that are *balls* which interact with active sites. The macroscopic behaviour of the ring derives from the individual (black or white) states of the balls. It is described by the colour ($\delta(t)$) whose calculation depends on the states of each ball.

In the representation underlying the *H*-theorem, the velocity of a particle depends on the number of collisions the particle had, while, in the Kac model, the colour of a ball depends on the number of times it left an active site.

Because the representation of the microdynamics matters above all, the particles need not to be faithfully represented in the model. Boltzmann's spheres share some well-known properties with the actual particles of gas (e.g. their spatial extent and their disposition of elastically colliding with each other), but the Kac's balls need not to possess such properties. Thus their only two possible states are "black" or "white".

The Kac ring also possesses additional features that are not contained in the representation underling the *H*-theorem, but are nevertheless harmless because they only interfere with the representational function of the model in irrelevant respects. For instance, their change of state depends on whether or not they have left an active site. Thus Kac could have used points or stars instead of active sites, or blue or gold balls, these fictional features would have had no influence for the purpose at stake.

5.2 Multi-realisability

A second important feature captured by the Kac ring that concerns the relation between the macroscopic and microscopic levels of the system is the multi-realisability of the macroscopic behaviour. In the idealised representation underlying the *H*-theorem, a state of the gas system, characterised by a certain *H*, can be obtained from different microscopic configurations. This is similar in the Kac ring. A state of the ring, characterised by a certain colour of the ring δ, can be obtained from different microscopic configurations of the ring.

An additional feature is the conservation of the number of the gas particles. In the Kac ring, the number of balls does not change and is therefore conserved.

5.3 The Molecular Chaos Hypothesis

A third important feature from the *H*-theorem is of course the molecular chaos hypothesis. It is not directly integrated into the model itself but constitutes an extrinsic assumption. In the *H*-theorem, the molecular chaos hypothesis corresponds to: $n_A = N.f(\mathbf{v};t).d\mathbf{v}$, $n_B = N.f(\mathbf{w};t).d\mathbf{w}$.

In the Kac model, the analog of the molecular chaos hypothesis consists in assuming that the active sites are randomly arranged and that the ratio of active sites equals $\frac{W_A(t)}{W(t)} = \frac{N_A}{N}$ and to $\frac{B_A(t)}{B(t)}$.

In the Kac ring, the randomness is in the placement of the active sites, not in the balls doing the motion ([26]). We should note that this idealisation helps the users in making inferential work. For example, collisions in the model occur always at the same locations, i.e. at the active sites. Thus the number and the location of collisions do not change over time. This has a cognitive virtue, as Bricmont highlights, insofar as "Scattering with fixed objects tends to be easier to analyse than collisions between particles" ([9], p. 157). This is also a harmless idealisation in that macroscopic irreversibility and microscopic reversibility are the most important aspects to be represented.

5.4 Reversibility and Recurrence

Lastly, the reversible feature of the microdynamics is also captured since, in the Kac model, the reversed behaviour of the ring is perfectly symmetric to its original behaviour. Furthermore the number of dimensions is reduced in the model. While the gas system is three-dimensional in Boltzmann's representation, the Kac's balls only can move along a (one-dimensional) ring. In this way, the user can easily see (concretely on a paper or a screen, or mentally) the reversibility of a change in the colour of a ball.

This reduction of dimensions also seems to enable recurrence. That the system is not idealised as a string for instance, but is rather represented as a ring, seems to enable recurrence in the system. If the system was idealised as a string, it would have been finite (since the system is supposed to be closed) and recursiveness would not have been possible. Since the ring is a one-dimensional closed space, the system can by contrast come back to its previous states.

Furthermore, in the model, the recurrence period strictly equals two periods of rotation. And yet, as actual gas systems have more degrees of freedom than the Kac ring, their recurrence periods are much longer than that. Nevertheless, in virtue of this idealisation, one is able to "observe" the phenomenon of recurrence, which would not have been possible some other way.

The recurrence time in the Kac ring is linear in N (i.e. the size of the system), whereas it is exponential in physical systems. But this idealisation is not essential once we recognise it.[6]

6 Conclusion

I have shown that the Kac model is used to draw lessons about Boltzmann's hypotheses. The model illustrates macroscopic irreversibility, enables us to derive the reversibility and recurrence paradoxes and to consider how they relate to Boltzmann's H-theorem.

This paper argues that all of this is possible because the Kac model distorts the (already idealised) representation of a gas system underlying the H-theorem, and

---

[6] I thank one of the reviewers for this comment.

highlights the most relevant aspects of the class of systems being studied. The idealisations it contains are active in the representing. They instantiate a relation between a macroscopic level and a microscopic level of a system, a multi-realisability of a macroscopic behaviour, an analog of the molecular chaos hypothesis, as well as a reversible and recurrent microscopic dynamics.

It may well be the case that, from a more general perspective, toy models contain idealisations that represent general features holding across different empirical target systems. That said, the Kac ring is a unique piece of scientific work that deserved a close analysis of its various idealisations.

References

1. Ardourel, V.: Irreversibility in the derivation of the boltzmann equation. Found. Phys. **47**(4), 471–489 (2017)
2. Batterman, R.W., Rice, C.C.: Minimal model explanations. Philos. Sci. **81**(3), 349–376 (2014)
3. Bokulich, A.: Can classical structures explain quantum phenomena? Br. J. Philos. Sci. **59**(2), 217–235 (2008)
4. Bokulich, A.: How scientific models can explain. Synthese **180**(1), 33–45 (2011)
5. Bokulich, A.: Explanatory models vs. predictive models: reduced complexity modeling in geomorphology. In: Recent Progress in Philosophy of Science: Perspective & Foundational Problems. Springer, Dordrecht (2013)
6. Boltzmann, L.: Über die beziehung zwischen dem zweiten haubtsatze der mechanischen wärmetheorie und der wahrscheinlichkeitsrechnung resp. dem sätzen über das wärmegleichgewicht wiener berichte. In: (Boltzmann 1909) vol. II, paper 42 **76**, 373–435 (1877)
7. Boltzmann, L.: Reply to zermelo's remarks on the theory of heat. Annalen der Physik **57**, 773–84 (1896)
8. Boltzmann, L.: On zermelo's paper: on the mechanical explanation of irreversible processes. Annalen der Physik **60**, 392–98 (1897)
9. Bricmont, J.: Science in chaos or chaos in science? Physicalia Mag. **17**, 159–208 (1995)
10. Coppersmith, M.: An example of the effect of irreversible microscopic dynamics on the macroscopic behavior of a large system. J. Stat. Phys. **20**(2), 175–179 (1979)
11. De Roeck, W., Jacobs, T., Maes, C.: An extension of the kac ring model. (2003). http://arxiv.org/pdf/cond-mat/0307460.pdf
12. Dresden, M.: New perspectives on kac ring models. J. Stat. Phys. **46**(5–6), 829–842 (1987)

13. Dresden, M., Feiock, F.: Models in nonequilibrium quantum statistical mechanics. J. Stat. Phys. **4**(2–3), 111–173 (1972)
14. Edens, B.: Semigroups and symmetry: an investigation of Prigogine's theories (2001)
15. Ehrenfest, P., Ehrenfest, T.: Über zwei bekannte einwände gegen das boltzmannsche h-theorem. Physikalische Zeitschrift **8**, 311–314 (1907)
16. Ehrenfest, P., Ehrenfest, T.: The Conceptual Foundations of the Statistical Approach in Mechanics. Dover Publications Inc, New York ([1912] 1990)
17. Elgin, C.: Telling instances. In: Frigg, R. (ed.) Beyond Mimesis and Convention. Springer, Dordrecht (2010)
18. Fernando, W.P.: Lack of molecular chaos and role of stochasticity in Kac's ring model. Master's dissertation, University of Akron (2009)
19. Gibbard, A., Varian, H.R.: Economic models. J. Philos. **75**(11), 664–677 (1978)
20. Giere, R.N.: How models are used to represent reality. Philos. Sci. **71**, 742–752 (2004)
21. Gottwald, G.A., Oliver, M.: Boltzmann's dilemma. An introduction to statistical mechanics via the kac ring. SIAM Rev. **51**, 613–635 (2009)
22. Grüne-Yanoff, T.: Appraising non-representational models. Philos. Sci. **80**(5), 850–861 (2013)
23. Hesse, M.B.: Models and Analogies in Science. University of Notre Dame Press, Notre Dame (1966)
24. Jebeile, J.: Explication et compréhension dans les sciences empiriques. Les modèles scientifiques et le tournant computationnel. PhD thesis, Université Paris 1 – Panthéon-Sorbonne (2013)
25. Jebeile, J., Kennedy, A.: Explaining with models : the role of idealizations. Int. Stud. Philos. Sci. **29**(4), 383–392 (2015)
26. Kac, M.: Some Stochastic Problems in Physics and Mathematics. Colloquium Lectures in Pure and Applied Science. Magnolia Petroleum CompanyMagnolia Petroleum CompanyMagnolia Petroleum CompanyMagnolia Petroleum CompanyMagnolia Petroleum CompanyMagnolia Petroleum CompanyMagnolia Petroleum Company. Field Research Laboratory, Dallas (1956)
27. Kac, M.: Probability and Related Topics in the Physical Sciences. Interscience Pub, New York (1959)
28. Kennedy, A.: A non representationalist view of model explanation. Stud. Hist. Philos. Sci. **43**(2), 326–332 (2012)
29. Knuuttila, T.: Modelling and representing: an artefactual approach to model-based representation. Stud. Hist. Philos. Sci. Part A **42**(2), 262–271 (2011)
30. Laymon, R.: Applying idealized scientific theories to engineering. Synthese **81**(3), 353–371 (1989)
31. Laymon, R.: Cartwright and the lying laws of physics. J. Philos. **86**(7), 353–372 (1989)
32. Luczak, J.: Talk about toy models. Stud. Hist. Philos. Sci. Part B **57**, 1–7 (2017)
33. Maxwell, J.C.: Tait's "thermodynamics". Nature **17**(431), 257–259 (1878)
34. McMullin, E.: Galilean idealization. Stud. Hist. Philos. Sci. Part A **16**(3), 247–273 (1985)
35. Nowak, L.: The idealizational approach to science: A survey. In: Brzezinski, J., Nowak, L. (eds.) Idealization III: Approximation and Truth. Poznán Studies in the Philosophy of the Sciences and the Humanities, vol. 25, pp. 9–63. Amsterdam and Atlanta, GA, Rodopi (1992)
36. Price, H.: Time's Arrow & Archimedes' Point. New Directions for the Physics of Time. Oxford University Press, Oxford (1996)
37. Schulman, L.S.: Time's Arrows and Quantum Measurement. Cambridge Univ. Press, Cambridge (1997)
38. Strevens, M.: Depth: An Account of Scientific Explanation. Harvard UP, Cambridge (2008)
39. Suárez, M.: An inferential conception of scientific representation. Philos. Sci. **71**, 767–779 (2004)
40. Suárez, M.: Scientific representation. Blackwell's philosophy. Compass **5**, 91–101 (2010)
41. Tavernier, J.: Classical and quantum descriptions of the Kac ring model. J. Stat. Phys. **14**(2), 101–128 (1975)
42. Weisberg, M.: Three kinds of idealization. J. Philos. **104**(12), 639–659 (2007)
43. Weisberg, M.: Who is a modeler? Br. J. Philos. Sci. **58**, 207–233 (2007)
44. Weisberg, M.: Simulation and Similarity: Using Models to Understand the World. Oxford University Press, Oxford (2013)