Relating the thermodynamic arrow of time to the causal arrow

Armen E Allahverdyan$^1$ and Dominik Janzing$^2$

$^1$ Yerevan Physics Institute, Alikhanian Brothers Street 2, Yerevan 375036, Armenia
$^2$ School of Electrical Engineering and Computer Science, University of Central Florida, Orlando, FL 32816-2362, USA
E-mail: aarmen@mail.yerphi.am and janzing@ira.uka.de

Received 8 August 2007
Accepted 27 February 2008
Published 4 April 2008

Abstract. Consider a Hamiltonian system that consists of a slow subsystem $S$ and a fast subsystem $F$. The autonomous dynamics of $S$ is driven by an effective Hamiltonian, but its thermodynamics is unexpected. We show that a well-defined thermodynamic arrow of time (second law) emerges for $S$ whenever there is a well-defined causal arrow from $S$ to $F$ and the back-action is negligible. This is because the back-action of $F$ on $S$ is described by a non-globally Hamiltonian Born–Oppenheimer term that violates the Liouville theorem, and makes the second law inapplicable to $S$. If $S$ and $F$ are mixing, under the causal arrow condition they are described by microcanonical distributions $P(S)$ and $P(S|F)$. Their structure supports a causal inference principle proposed recently in machine learning.

Keywords: new applications of statistical mechanics

©2008 IOP Publishing Ltd and SISSA 1742-5468/08/P04001+21$30.00
1. Introduction

In this paper we establish a relation between the causal arrow—i.e., emergence of a unidirectional interaction between two interacting systems—and the thermodynamic arrow of time. Studying causation in the context of various physical arrows of time is not a new subject [1]–[3]. One of the motivations for these studies is the analogy between the temporal asymmetry implied by the thermodynamic arrow and the asymmetry between the cause and effect: causes influence their effect, but not vice versa, and causes can only happen before their effects [1]–[3].

Causal explanations in everyday life often construct causal structures among phenomena that are not well localized in time (e.g., if one studies relations between crime and poverty in social sciences). Even for such phenomena we observe sometimes well-defined causal connections where one phenomenon is the cause and another one the effect. For understanding the link between thermodynamics and causality within a statistical
Relating the thermodynamic arrow of time to the causal arrow

physics setting, it is helpful to study the conditions under which we can consider one of two interacting systems as the cause and the other as the effect. The question is then to what extent the unidirectionality of the influence is related to the thermodynamics of the two systems.

The results presented provide some answers to the above general question. For describing those answers we proceed with separate introductions on the thermodynamic arrow and the causal arrow. This section then closes with qualitative discussions of our main results.

1.1. The thermodynamic arrow of time

Thermodynamic arrow of time refers to formulations of the second law. The understanding of this law from the first principles of quantum or classical dynamics is achieved within statistical physics (in contrast to thermodynamics, where the second law is postulated). In this statistical physics context we list the following basic formulations of the second law [4,7,8].

- Entropy formulation: coarse-grained entropy does not decrease in time for a closed system that starts to evolve from a certain non-equilibrium state [4]–[7].
- Thomson’s formulation: for a system that starts to evolve from an equilibrium state, no work extraction is possible by means of a cyclic process driven by an external source of work [7,8].

These statements entail an arrow of time, since they refer to the difference between final and initial values of the entropy and energy\(^3\), respectively. Each formulation has two different aspects: special initial conditions (non-equilibrium states for the entropy formulation, equilibrium states for Thomson’s formulation) and specific dynamic features of the system (closed dynamics, cyclic processes). Both of these aspects were studied from first principles [1,2], [4]–[9], [11,14,15]\(^4\).

There are more formulations of the second law, such as the minimal work principle [4,11,12] and the Clausius formulation [4,9,10]. These formulations are equivalent within phenomenological thermodynamics, but their consistent derivation from first principles shows that this equivalence does not always hold [10,11]. The Thomson and entropy formulations do not require anything more than a Hamiltonian dynamics that satisfies the Liouville equation [4,7,8], while the minimum work principle and the Clausius formulation do have additional requirements: an ergodic observable of work for the minimum work principle [11] and a weak (or conserved) interaction Hamiltonian for the Clausius formulation [4,9,10].

We shall thus focus on the Thomson and entropic formulations. Here the preference should be given to Thomson’s formulation, since there is no universally accepted definition of physical entropy for non-equilibrium states. In contrast, there is such a definition for

---

\(^3\) Since any interaction with an external source of work can be seen as a thermally isolated process, work is a difference between average energies; see section 6 for details.

\(^4\) The fact that we impose initial, and not final, conditions cannot be derived from the first principles. Instead, it should be taken as a fact that experiments are described by their initial conditions rather than being described by the final conditions. This fact alone does not ensure the existence of the thermodynamic arrow of time. The latter requires special initial conditions and specific dynamic mechanisms.

doi:10.1088/1742-5468/2008/04/P04001
work \cite{4,7}. The formulation and derivation of the entropy and Thomson’s formulation will be recalled below in section 6.

1.2. The causal arrow

The causal arrow refers to a dynamical situation when one degree of freedom \( C \) (the cause) influences another \( E \) (the effect), but does not get a back-reaction\(^5\). In this context we shall recall two operational definitions of the causal arrow. (i) Cutting off the interaction between \( C \) and \( E \) does not alter the dynamics of \( C \), while it influences the dynamics of \( E \). (ii) Perturbing the dynamics of \( C \) — e.g., by means of external fields, or by changing the initial conditions of \( C \) — will influence the dynamics of \( E \), while perturbing the dynamics of \( E \) will not influence \( C \).

In studying causal relations between general variables — that do not necessarily represent degrees of freedom of physical systems (e.g. in economy, medicine, social sciences) — scientific reasoning often depends on statistical data that have been obtained from mere observations. This is because interventions that would prove causal relations are often impossible. One then tries to draw plausible causal conclusions merely from stochastic dependences in the joint distribution function \( P(C, E) \) \cite{17}. In spite of their obvious importance—as sometimes our very survival depends on the proper identification of the cause versus effect—such conclusions cannot be always correct; they are merely plausible in the sense that they lead to correct predictions more frequently than they fail\(^6\).

Several criteria are known for this type of causal reasoning, if the number \( n \) of variables is 3 or more \cite{17}. In the case of two variables \( X_1, X_2 \) the task of distinguishing between ‘\( X_1 \) causes \( X_2 \)’ and ‘\( X_2 \) causes \( X_1 \)’ (given that exactly one of the statements is true) is a difficult one and there is no generally accepted approach to its solution. However, \cite{18}–\cite{20} present a first approach to addressing this task.

1.3. Purposes and results of the present work

(1) We shall follow in detail how the causal arrow and the thermodynamic arrow of time emerge in a closed, classical Hamiltonian system that consists of two subsystems \( S \) and \( F \). For studying the causal arrow it is natural to assume that \( S \) is slow, while \( F \) is fast.

In a more general perspective, the idea of slow versus fast variables has been developed in several different contexts, e.g., the slaving principle proposed by Haken as a cornerstone for synergetics, self-organization, and hierarchical dynamics \cite{22}. Indeed, many (almost all?) models studied in mechanics, (non-)equilibrium statistical physics, chemical kinetics, mathematical ecology, etc, are not fundamental, but rather describe the effective behavior of slow degrees of freedom.

(2) The absence of the causal arrow in the above closed system is quantified by the back-reaction of \( F \) on \( S \). Under some natural conditions outlined below, this back-reaction amounts to an additional (Born–Oppenheimer\(^7\)) term in the Hamiltonian of \( S \).

\(^5\) By the causal arrow we thus do not mean the macroscopic causality that occurs when the past macrostate determines the future one.

\(^6\) The fact that stochastic dependences cannot serve as the basis for drawing unique causal conclusions was stressed by Hume \cite{16}.

\(^7\) The names come from the early days of atomic physics, when M Born and R Oppenheimer calculated in the quantum mechanical setting the force exerted by fast electrons on slow nuclei.
The dynamics of $S$ is then autonomous and energy conserving. However, the Born–Oppenheimer term has the following peculiar feature: it depends explicitly on the initial value of the coordinates of $S$ that participate in the interaction with $F$. This is a consequence of memory generated during the tracing out of the fast variables. Thus there is no single, global Hamiltonian for $S$. We shall show that due to this fact the basic formulations of the second law—and thus the usual formulation of the thermodynamic arrow of time—do not apply to $S$, even if we assume the existence of proper initial conditions\(^8\).

(3) If the Born–Oppenheimer term can be neglected for the dynamics of $S$, the applicability of the second law for $S$ is recovered. This neglect indicates the existence of the causal arrow in the system considered: $S$ appears to be the cause for $F$. Thus the local thermodynamic arrow for $S$ emerges due to the causal arrow.

Note that the second law applies to the fast subsystem $F$, which has a driven, globally Hamiltonian dynamics. Such a dynamics serves as a basis for deriving the second law from first principles [4], [7]–[9], [11,12].

(4) Another important consequence of the Born–Oppenheimer term is that it can make $S$ non-ergodic, even if the bare Hamiltonian of $S$ is assumed to have ergodic features. (For the definition employed or ergodicity see the discussion around (9), (10); for the precise definition of what we mean by non-ergodicity see the discussion around (21).) Indeed, no microcanonical distribution can be introduced for $S$, unless the Born–Oppenheimer term is neglected. We show that together with the emergence of the causal arrow, there appear natural, microcanonical probability distributions\(^9\) $P(S)$ and $P(F|S)$, where $P(S)$ and $P(F|S)$ are simpler (in the precise sense discussed below) than, respectively, $P(F)$ and $P(S|F)$.

In section 2 we define the system to be studied. Sections 3 and 4 discuss, respectively, the dynamics of the fast subsystem $F$ and the convergence of its probability distribution toward the microcanonical distribution. The dynamics of the slow subsystem $S$ is described in section 5. In section 6 we discuss in detail the (in)applicability of the basic statements of the second law (thermodynamic arrow) to the dynamics of $S$. The joint emergence of the thermodynamic arrow and the causal arrow is outlined in section 7. Section 8 relates the results obtained to the simplicity principle proposed recently in machine learning. The last section presents our conclusions and offers some speculations.

2. Fast and slow subsystems

While the above discussion on the causal arrow and the second law has been rather general, here we study a concrete model for coupled fast ($F$) and slow ($S$) classical degrees of freedom.

The overall Hamiltonian of $S + F$ is defined as

$$\mathcal{H}(\Pi, Q, z) = H_s(\Pi, Q) + H_f(z) + H_I(z, Q),$$

where $z = (q_1, \ldots, q_N; p_1, \ldots, p_N)$ are canonical coordinates and momenta of $F$, and where $Q = (Q_1, \ldots, Q_M)$ and $\Pi = (\Pi_1, \ldots, \Pi_M)$ are, respectively, canonical coordinates and

---

\(^8\) This does not mean that there cannot be other—apart from the thermodynamic arrow in the sense explained in the introduction—temporal asymmetries in the dynamics of $S$.

\(^9\) $P(F|S)$ is the conditional probability for the coordinates and momenta of $F$, with the variables of $S$ being fixed.
Relating the thermodynamic arrow of time to the causal arrow

momenta of $S$. The bare Hamiltonian of $S$ is $H_s(\Pi, Q)$, while $H_f(z)$ and $H_I(z,Q)$ are, respectively, the bare Hamiltonian of $F$ and the Hamiltonian for the interaction between $S$ and $F$. The interaction involves only the coordinates $Q$ and $z$ and not the momentum of the slow system. We adopt this physically natural assumption for having better control in the division between the fast and slow variables; see below. We shall join $H_f(z)$ and $H_I(z,Q)$ into a single expression $H(z,Q)$, since we show below, and this is well known in general, that provided there is a well-defined timescale separation between $S$ and $F$, $H(z,Q)$ plays the role of the fast Hamiltonian. Thus the overall Hamiltonian reads

$$H(\Pi, Q, z) = H_s(\Pi, Q) + H(z, Q).$$

(2)

Let $\tau_f$ be the characteristic time of $F$ for the slow variable $Q$ being fixed (for a more precise definition see after (9)). We shall assume that both $Q$ and $\dot{Q}$ are slow variables with respect to $\tau_f$. This assumption is consistent with the fact that the $S$–$F$ coupling involves only the coordinate $Q$ of $S$: according to the Hamiltonian equation, $\dot{Q} = \partial_{\Pi}[H_s(\Pi, Q)]$, generated by (2), $\dot{Q}$ does not depend explicitly on the fast variable $z$. (Indeed, were $\dot{Q}$ to depend on $z$ explicitly, its derivative $\ddot{Q}$ would be large, and thus $\dot{Q}$ would not be slow.)

Define $\nu_Q$ and $\nu_{\dot{Q}}$ as the characteristic times over which $Q$ and $\dot{Q}$ change. Define

$$\tau_Q \equiv \min(\nu_Q, \nu_{\dot{Q}}).$$

(3)

Thus our basic assumption on the separated timescales (adiabatic limit) reads

$$\tau_f \ll \tau_Q.$$  

(4)

3. Energy of the fast subsystem

Our intention is to see how the energy $H(z, Q)$ of the fast subsystem $F$ changes in time.

Hamilton’s equations of motion for the fast subsystem imply $(d/dt)H(z_t, Q_t) = \dot{Q}_t \partial_Q H(z_t, Q_t)$. Assuming the adiabatic limit $\tau_f \ll \tau_Q$, and denoting as $Q_t$ and $z_t$ the time-dependent coordinates, we have for the energy change for the intermediate times $\tau_Q \gg \tau \gg \tau_f$

$$\frac{d}{d\tau} E \equiv \frac{1}{\tau}[H(z_{\tau+\tau}, Q_{\tau+\tau}) - H(z_\tau, Q_\tau)]$$

$$= \int_{\tau}^{\tau+\tau} ds \frac{dH}{ds}(z_s, Q_s)$$

$$= \frac{\dot{Q}_t}{\tau} \int_{\tau}^{\tau+\tau} ds \partial_Q H(z_s, Q_t) + o\left(\frac{\tau}{\tau_Q}\right),$$

(5)

where we took $\dot{Q}_t$ out of the integral, since $\dot{Q}_t$ (together with $Q_t$) is assumed to be a slow variable.

doi:10.1088/1742-5468/2008/04/P04001
The last integral in (5) refers to the $Q = \text{const}$ dynamics with $Q_t = Q$. This dynamics has a constant energy $E = H(z, Q_t)$. Define for the microcanonical distribution

$$\mathcal{M}(z, E, Q) \equiv \frac{1}{\omega(E, Q)} \delta[E - H(z, Q)],$$

(6)

$$\omega(E, Q) \equiv \int dz \delta[E - H(z, Q)],$$

(7)

where $\omega(E, Q)$ ensures the proper normalization: $\int dz \mathcal{M}(z, E, Q) = 1$.

Consider the following obvious relation:

$$\int dz w(z) \mathcal{M}(z, E) = \frac{1}{\tau} \int_{t}^{t+\tau} ds \int dz w(z) \mathcal{M}(z, E),$$

(8)

where $w(z) \equiv \partial_Q H(z, Q_t)$, and where for simplicity we drop the explicit dependence on $Q = Q_t = \text{const}$.

In the RHS of (8) we change the integration variable using $y = T_t^{-s} z$, where $T_t$ is the flow generated by the Hamiltonian $H(z) = H(z, Q_t)$ between times 0 and $t$. Employing Liouville’s theorem, $dz = dy$, and energy conservation, $\mathcal{M}(z, E) = \mathcal{M}(y, E)$, one gets

$$\int dy \mathcal{M}(y, E) \frac{1}{\tau} \int_{t}^{t+\tau} ds \ w(T_{s-t} y).$$

(9)

If $w(z)$ is an ergodic observable of the $Q_t = \text{const}$ dynamics, then by the definition of ergodicity there is a characteristic time $\tau_f$ such that for $\tau \gg \tau_f$ the time average in (9) depends on the initial condition $y$ only via its energy $H(y, Q_t)$ [23, 25]. Since $\mathcal{M}(y, E)$ is proportional to a $\delta$-function at $E = H(z, Q)$, the integration over $y$ in (9) drops out, and we get that the time average in (8) is equal to the microcanonical average at the energy $E$. Applying this to the time average in (5) we get

$$\frac{dE}{d\tau} = \frac{dQ}{d\tau} \int dz \partial_Q H(z, Q_t) \mathcal{M}(z, E_t, Q_t),$$

(10)

where we noted again that $\dot{Q}$ is a slow variable.

We define the phase-space volume $\Omega$ enclosed by the energy shell $E$:

$$\Omega(E, Q) \equiv \int dz \theta(E - H(z, Q)).$$

(11)

Let us see how $\Omega(E, Q)$ changes in the slow time:

$$\frac{d}{d\tau} \Omega(E, Q) = \partial_E \Omega|_Q \frac{dE}{d\tau} + \partial_Q \Omega|_E \frac{dQ}{d\tau}.$$  

(12)

Using (10), (11) we get

$$\partial_Q \Omega|_E \frac{d}{d\tau} \Omega|_Q = - \int dz \partial_Q H(z, Q_t) \mathcal{M}(z, E_t, Q_t),$$

(13)

and then from (10), (12), (13),

$$\frac{d}{d\tau} \Omega(E, Q) = \partial_E \Omega|_Q \left[ \frac{dE}{d\tau} + \frac{dQ}{d\tau} \partial_E \Omega|_Q \right]$$

$$= \partial_E \Omega|_Q \left[ \frac{dE}{d\tau} - \frac{dE'}{d\tau} \right] = 0.$$  

(14)

doi:10.1088/1742-5468/2008/04/P04001
Thus, the phase-space volume $\Omega(E, Q)$ is an adiabatic invariant, i.e., it is conserved within the slow dynamics. In particular, in the adiabatic limit the points of the fast phase space located initially at the energy shell $E_i$ appear on the energy shell $E_f$, which is found from

$$\Omega(E_i, Q_i) = \Omega(E_f, Q_f).$$

Since by definition (11), $\Omega(E)$ is an increasing function of $E$, for given $Q_i, Q_f$ and $E_i$, equation (15) has a unique solution

$$E_f \equiv h(Q_f | E_i, Q_i).$$

In the adiabatic limit the energy $h(Q_f | E_i, Q_i)$ of $F$ does not depend on the precise phase-space location of the fast trajectory on the energy shell $E_i$.

Note that the derivation of (12) does not demand full ergodicity—which means that all smooth observables of $F$ are ergodic—only a certain observable is assumed to be ergodic [23]. The argument expressed by (8), (9) applies to calculating the time average of any ergodic observable $w(z)$ of $F$ for a fixed $Q$.

The adiabatic invariance of $\Omega$ for ergodic systems is well known [24]–[26] and motivated the microcanonical definition of entropy as $\ln \Omega$ [25,26]. The precision of the invariance is studied in [27]. We presented the above derivation for completeness of this work and to highlight the two basic assumptions that are not properly articulated in the literature: (i) ergodicity of an observable versus the full ergodicity, and (ii) the necessity for both $Q$ and $\dot{Q}$ being slow.

4. Conditional microcanonical distribution of the fast subsystem

For describing time averages of ergodic observables of $F$ (see (8), (9) and the discussion after (10)) we can employ the following time-dependent microcanonical conditional probability:

$$P_f[z|Q_i, \Pi_i] = \frac{\delta[h(Q_\tau | E_i, Q_i) - H(z, Q_\tau)]}{\int dz \delta[h(Q_\tau | E_i, Q_i) - H(z, Q_\tau)].}$$

Below we explain how to find $Q_\tau$ given the initial energy $E_i$ of $F$, the initial canonical coordinates $Q_i, \Pi_i$ of $S$ and the time $\tau$. Note that $P_f[z|Q_i, \Pi_i]$ is time dependent and varies with time on the slow timescale $\tau \sim \tau_Q$.

There is another way of introducing the microcanonical distribution (17) which explicitly uses the ensemble description [28,29]. If for a fixed $Q$ the system $F$ is mixing, then for any sufficiently smooth initial probability distribution $p(z, 0)$ of $F$, the ensemble averages of sufficiently smooth (i.e., sufficiently coarse-grained) observables $A(z)$ of $F$ converge in time to the averages taken over the (17) [28,29]:

$$\int dz \ p(z, t) A(z) \to \int dz \ P_f[z|Q_i, \Pi_i]\ A(z).$$

The rate of this convergence defines the mixing time. It is more natural (especially for chaotic systems) to define observables via ensemble averages than via averages over time [28]. If not stated otherwise, from now on we assume that $F$ is mixing, and thus the mixing time coincides with $\tau_f (\ll \tau_Q)$ defined around (9). For strongly
Relating the thermodynamic arrow of time to the causal arrow

(and homogeneously) chaotic systems the mixing time is inversely proportional to the KS entropy [28, 29].

5. Dynamics of the slow subsystem

Let us average the equations of motion \( \dot{\Pi} = -\partial_Q[H_s(\Pi, Q) + H(Q, z)] \) and \( \dot{Q} = \partial_\Pi[H_s(\Pi, Q)] \) over the microcanonical distribution (17). We get that S is by itself a Hamiltonian system:

\[
\frac{d}{d\tau} \Pi = -\partial_Q H_s, \tag{19}
\]

\[
\frac{d}{d\tau} Q = \partial_\Pi H_s, \tag{20}
\]

with an effective Hamiltonian

\[
H_s(\Pi, Q|Q_i, E_i) = H_s(\Pi, Q) + h(Q|Q_i, E_i), \tag{21}
\]

which is the sum of \( H_s(\Pi, Q) \) and the Born–Oppenheimer term \( h(Q|Q_i, E_i) \). In particular, \( H_s(\Pi, Q|Q_i, E_i) \) determines the actual slow trajectory \( Q_\tau \), given its initial location \( (\Pi_i, Q_i) \). Substituting this back into (17) we thus complete the description of F.

The evolution generated by (19) conserves the energy \( H_s \). This is the total energy of S+F. Note that the Born–Oppenheimer term \( h(Q|Q_i, E_i) \) depends on the initial coordinate \( Q_i \). This means that the points in the phase space \( (\Pi, Q) \) that had initially equal energy (but different initial coordinates \( Q_i \)) will have different energies at later times. Thus S is not globally Hamiltonian.

While this fact seems to be of no special importance when we consider a single slow trajectory, it matters greatly for developing statistical physics for S. Indeed, there is no global slicing of the phase space into energy shells which makes the definition of the microcanonical distributions impossible.

Thus S is non-ergodic: once ergodic systems are characterized by losing the memory on the initial phase-space location and remembering only the initial energy (recall the argument around (8), (9)), in the situation considered the very form of the energy depends on the initial phase-space location.

5.1. Liouville equation and Liouville theorem

A consequence of the non-globally Hamiltonian dynamics is that the Liouville equation and the corresponding theorem do not hold. With the Hamilton equations (19) one can relate a conditional probability

\[
P_{\text{con}}(\Pi, Q, \tau|\Pi_i, Q_i, 0) = \delta(\Pi - \Pi(\Pi_i, Q_i, \tau))\delta(Q - Q(\Pi_i, Q_i, \tau)), \tag{22}
\]

where \( \Pi(\Pi_i, Q_i, \tau) \) and \( Q(\Pi_i, Q_i, \tau) \) are the solutions of (19), with initial conditions \( (\Pi_i, Q_i) \).

As follows from (19), (22), \( P_{\text{con}}(\Pi, Q, \tau|\Pi_i, Q_i, 0) \) does satisfy the Liouville equation

\[
\partial_\tau P_{\text{con}} = \partial_Q H_s \partial_\Pi P_{\text{con}} - \partial_\Pi H_s \partial_Q P_{\text{con}}. \tag{23}
\]

Were \( H_s \) not dependent on \( Q_i \), the direct integration of (23) with the initial distribution \( P(\Pi_i, Q_i, 0) \) would produce the Liouville equation for the unconditional probability.
\( \mathcal{P}(\Pi, Q, \tau) \). But since \( \mathcal{H}_s(\Pi, Q|Q_i) \) does depend on \( Q_i \), the integration with \( \mathcal{P}(\Pi_i, Q_i, 0) \) does not lead to a Liouville equation for \( \mathcal{P}(\Pi, Q, t) \).

Thus the Liouville equation and together with it the Liouville theorem (conservation of the phase-space volume) do not hold. Below we shall demonstrate this on an explicit example.

### 5.2. An example

We assume that F and S without mutual coupling are two free particles, with masses \( m \) and \( M \), respectively. The S–F coupling creates a harmonic potential for F:

\[
H(p, q, Q) = \frac{p^2}{2m} + \frac{Q^2q^2}{2}.
\]  

(24)

If we regard the slow variable \( Q \) as a parameter, F is an ergodic system with the characteristic time

\[
\tau_f = \frac{2\pi\sqrt{m}}{Q}.
\]  

(25)

Equation (12) reduces to the conservation of action: \( E/|Q| = \text{const.} \) and thus the Born–Oppenheimer potential \( h(Q|E_i, Q_i) \) reads from (15)

\[
h(Q|E_i, Q_i) = E_i\frac{|Q|}{|Q_i|}.
\]  

(26)

As the simplest example of a bare slow Hamiltonian we can take free motion with a mass \( M \):

\[
H_s = \frac{\Pi^2}{2M}.
\]  

(27)

Thus the dynamics of the slow subsystem S is described by the effective Hamiltonian:

\[
\mathcal{H}_s = \Pi^2/2M + E_i|Q|/|Q_i|.
\]

Assume that \( Q > 0 \) and solve the Hamilton equations as

\[
\Pi(\tau) = \Pi_i - \frac{E_i\tau}{Q_i}, \quad Q(\tau) = -\frac{E_i\tau^2}{2MQ_i} + \frac{\Pi_i\tau}{M} + Q_i,
\]  

(28)

where the initial time was taken as \( \tau = 0 \). The characteristic time \( \nu_Q \) of \( Q \) can be estimated from \( Q(\nu_Q) - Q_i \sim Q_i \):

\[
\nu_Q = \min \left[ \frac{MQ_i}{\Pi_i}, \sqrt{\frac{2MQ_i^2}{E_i}} \right].
\]  

(29)

For the characteristic time \( \nu_\dot{Q} \) of \( \dot{Q} \) (estimated via \( \dot{Q}(\nu_\dot{Q}) - \dot{Q}_i \sim \dot{Q}_i \)) we get

\[
\nu_\dot{Q} = Q_i\Pi_i/E_i.
\]  

(30)

If \( \Pi_i \to 0 \) we should take \( \nu_\dot{Q} = \sqrt{2MQ_i^2/E_i} \).

It is seen now that unless \( Q(\tau) \simeq 0 \), the adiabatic conditions \( \nu_Q \gg \tau_f \) and \( \nu_\dot{Q} \gg \tau_f \) can be satisfied, e.g., for a sufficiently small \( m \) and sufficiently large \( M \).
Relating the thermodynamic arrow of time to the causal arrow

One now has from (28) for the determinant of the Jacobian matrix

\[ J(\tau) \equiv \frac{\partial (\Pi(\tau), Q(\tau))}{\partial (\Pi_i, Q_i)} = 1 - \frac{E_i \tau^2}{2M Q_i^2}, \]  

(31)

which is not equal to 1. Moreover, its absolute value can be either larger or smaller than 1, since it is not difficult to see that the conditions \( Q > 0 \) and \( E_i \tau^2/2M Q_i^2 > 2 \) can be satisfied together.

Perhaps the most visible consequence of the absence of the Liouville theorem is that the fine-grained entropy

\[ S_{fg}[\tau] = -\int d\Pi dQ \mathcal{P}(\Pi, Q, \tau) \ln \mathcal{P}(\Pi, Q, \tau), \]  

(32)

of the slow subsystem is not constant any longer. Indeed, take a small phase-space volume \( v(0) \) and assume that \( \mathcal{P}(\Pi, Q, 0) \) is constant inside this volume and equal to zero outside. The fine-grained entropy (32) is then \( S_{fg}[\tau] = \ln v(\tau) \), where \( v(\tau) \) is obtained from \( v(0) \) under the action of the flow generated by Hamiltonian \( \mathcal{H}_s \). Thus, \( S_{fg}[\tau] - S_{fg}[0] = \ln v(\tau)/v(0) = \ln |J(\tau)| \) can both increase and decrease over the course of time, as (31) illustrates.

When can one neglect the non-conservation of the phase-space volume? Taking in (31) \( \tau \sim \tau[Q] \), and going in (29) to the limit of a small \( E_i \) or a large \( M \), we get that the non-conservation of the phase-space volume can be neglected—though \( Q \) still changes significantly—if the fast energy \( E_i \) is much smaller than the bare slow energy \( \Pi_i^2/2M \).

6. Thermodynamic arrow for the slow subsystem

6.1. Thomson formulation of the second law

How does the second law apply to the effectively Hamiltonian, autonomous slow subsystem \( S \)? The basic formulation of the second law is due to Thomson: no work can be extracted from an initially equilibrium system via a cyclic change of an external field. This statement is derived as a theorem both in classical and in quantum mechanics [7, 8]. We already argued as to why this formulation is superior to the entropy formulation: entropy is not directly observable and there is no general consensus on the definition of thermodynamic entropy for a non-equilibrium state. In contrast, work is directly observable, has a clear mechanical meaning, and its general definition is universally accepted [4, 7]. Here we focus on Thomson’s formulation, while the entropic formulation is studied below.

Let us recall the statement of the Thomson formulation when no interaction between \( S \) and \( F \) is present, i.e., the dynamics of \( S \) is generated by

\[ \mathcal{H}_s(\Gamma, \lambda_\tau), \quad \Gamma \equiv (Q, \Pi). \]  

(33)

The interaction of \( S \) with an external source of work is described by a time-dependent field \( \lambda_\tau \) [4, 7].

Let the initial phase-space points be sampled according to the Gibbs distribution:

\[ \mathcal{P}_G(\Gamma) = \frac{e^{-\beta \mathcal{H}_s(\Gamma)}}{Z}, \quad Z = \int d\Gamma e^{-\beta \mathcal{H}_s(\Gamma)}, \]  

(34)

doi:10.1088/1742-5468/2008/04/P04001
where $\beta = 1/T > 0$ is the inverse temperature. A cyclic change of the external field means

$$\lambda_0 = \lambda_{\tau_c} = \lambda,$$

(35)

where $\tau_c$ is the cycle time.

For the thermally isolated process considered the work is defined as the average energy difference\(^{10}\), and the statement of the Thomson formulation reads [7, 8]:

$$W = \int d\Gamma H_s(\Gamma, \lambda)[\tilde{P}(\Gamma, \tau_c) - P_G(\Gamma)] \geq 0,$$

(36)

where $\tilde{P}(\Gamma, \tau_c)$ is the final (at $t = \tau_c$) probability distribution obtained from the initial Gibbsian probability distribution $P_G(\Gamma)$ via the Liouville equation with the time-dependent Hamiltonian (33).

The inequality in (36) is based on the following three facts: (i) initial and final Hamiltonians are the same due to (33), (35); (ii) the same Hamiltonian appears in the initial Gibbs distribution; (iii) the Liouville equation.

The easiest way to establish the validity of (36) is to employ the positivity of the relative entropy [7]:

$$S_{\tilde{P}(\tau_c) \parallel P_G} \equiv \int d\Gamma \tilde{P}(\Gamma, \tau_c) \ln \frac{\tilde{P}(\Gamma, \tau_c)}{P_G(\Gamma)} \geq 0,$$

(37)

which holds for any probability distributions $\tilde{P}(\Gamma, \tau_c)$ and $P_G(\Gamma)$. Employing in (37) the conservation of the fine-grained entropy, $S_{\text{fg}}[\tilde{P}(\tau_c)] = S_{\text{fg}}[P_G]$, due to the Liouville theorem, we get

$$S_{\text{fg}}[P_G(\Gamma)] = S_{\text{fg}}[\tilde{P}(\Gamma, \tau_c)] \geq 0,$$

(38)

and then substituting (34) into $\ln P_G(\Gamma)$ in (38) and recalling (35) we arrive at (36).

Let us now return to the slow subsystem $S$ coupled to $F$. Now the slow Hamiltonian is given by (21) instead of (33). At the initial time, these two Hamiltonians are equal modulo a factor $E_i$. We shall assume that the initial probability for $\Pi$ and $Q$ is still given by (34), while initially the fast system always starts with the same energy $E_i$. For instance it is described by the microcanonical probability distribution (17), and then the overall initial distribution of $S$ and $F$ is the product of the above specified marginal distributions for $S$ and $F$.

Thus the overall distribution is not Gibbsian and the applicability of the Thomson formulation to the overall system is not automatic. The work is still given by the average energy difference (of the slow subsystem, or, equivalently, of the total system) calculated via the effective slow Hamiltonian (21). This can be argued for in exactly the same way.

---

\(^{10}\) Work for a single trajectory $(\Pi_\tau, Q_\tau)$ is defined as $W = \int_0^\tau du \partial_{\lambda_u} H_s(\Pi_u, Q_u, \lambda_u) d\lambda_u/du$. Employing the Hamilton equations of motion we get $W = H_s(\Pi_\tau, Q_\tau, \lambda_\tau) - H_s(\Pi_i, Q_i, \lambda_i)$, where $(\Pi_\tau, Q_\tau, \lambda_\tau)$ and $(\Pi_i, Q_i, \lambda_i)$ are the corresponding initial and final values. Averaging this expression over the initial and final values, and recalling (35), we get the expression for work as the average energy difference (36).
where the potential \( V \) comes from (34). The first term \( TS[\mathcal{P}(\tau_c) \parallel \mathcal{P}_G] \) in the RHS of (40) is non-negative. The fine-grained entropy difference \( S_{fg}[\mathcal{P}(\tau_c)] - S_{fg}[\mathcal{P}_G] \) does not have definite sign, since the Liouville equation does not hold. Moreover, the RHS of (40), equal to \( T \int d\Gamma [\mathcal{P}_G(\Gamma) - \mathcal{P}(\Gamma, \tau_c)] \ln \mathcal{P}_G(\Gamma), \) does not have a definite sign either. Even if the latter term is positive — e.g., because the fine-grained entropy increased in time: \( S_{fg}[\mathcal{P}(\tau_c)] > S_{fg}[\mathcal{P}_G] \) — the term in (40) does not have any reason to be positive. Apart from special coincidences, there is no reason why the two ‘dangerous’ terms \( S_{fg}[\mathcal{P}(\tau_c)] - S_{fg}[\mathcal{P}_G] \) and (40) would cancel each other.

Thus the proof of Thomson’s formulation can fail twice: once because the Liouville equation does not hold, and a second time because a cyclic change (35) of the parameter \( \lambda \) does not yet imply a cyclic change of the Born–Oppenheimer term (this is the origin of the term in (40)).

The latter aspect can be studied separately. Let \( S \) be a single particle, and assume the following natural choice of the bare slow Hamiltonian: \( H_s(\Pi, Q) = \Pi^2/2M + V(Q) \), where the potential \( V(Q) \) has its deepest minimum at \( Q_0 \): \( V(Q) > V(Q_0) \) for \( Q \neq Q_0 \). In the initial Gibbs distribution of \( S \) take \( T = 0 \). Then the initial distribution is reduced to a single initial condition \( \Pi_0 = 0 \), and \( Q_0 = Q_0 \). The interaction of \( S \) with external sources of work is described by an additional potential \( u(Q, \lambda_r) \), which is equal to zero both initially and at the end of the cycle; see (35). We assume that at intermediate times, \( u(Q, \lambda_r) \) is such that \( Q_0 \) ceases to be a local minimum of the overall potential, i.e., the particle located initially at \( Q_0 \) will move out of it and will change its energy. Now for the work one has analogously to (40)

\[
W = H_s(\Pi(\tau_c), Q(\tau_c)) - H_s(0, Q_0) + h(Q(\tau_c)|E_i, Q_0) - E_i, \tag{41}
\]

where \( \Pi(\tau_c) \) and \( Q(\tau_c) \) are the values of the canonical coordinates at the end of the cyclic process. They are obtained from solving (19), (21). The term in (41) corresponds to that in (40).

While \( H_s(\Pi(\tau_c), Q(\tau_c)) - H_s(0, Q_0) \) is non-negative by construction, there is no general restriction on the sign of \( h(Q(\tau_c)|E_i, Q_0) - E_i \). Noting the freedom in choosing \( h(Q|E_i, Q_i) \), one can make \( h(Q(\tau_c)|E_i, Q_0) - E_i \) so negative that the overall work is negative as well:
$W < 0$. Here is a concrete numerical example for this effect. Let us take

$$H_s(\Pi, Q) = \frac{\Pi^2}{2M} + \frac{a}{2}(Q - Q_0)^2 + Qf(t), \quad (42)$$

$$h(Q|E_i, Q_0) = \frac{E_i|Q|}{|Q_0|}, \quad (43)$$

that is for the bare slow motion we take a harmonic oscillator, while the Born–Oppenheimer term is adopted from our discussion in (26). In (42) $a > 0$ stands for the strength of the harmonic potential, while $f(t)$ is an external field acting on the slow particle. Equations (42), (43) generate the following equation of motion for the slow coordinate $Q$:

$$M\ddot{Q} = -a(Q - Q_0) - f(t) - E_i\text{sign}(Q)|Q_0|, \quad (44)$$

which is to be solved with the initial conditions $\dot{Q}(0) = 0$ and $Q(0) = Q_0$. For the parameters involved we take: $M = Q_0 = E_i = 1$, $a = 0.01$. We also assume $f(t) = 2\sin(2t)$ for the external field. It was checked numerically that at $t = \pi/2$, where the external field has changed cyclically, the work defined by (41) is negative: $W = -0.2448$.

### 6.2. Entropic formulation of the second law

The invalidity of the entropic formulation is studied along similar lines. Assume that $S$ consists of several subsystems: $(\Pi; Q) = (\Pi_1, \ldots, \Pi_M; Q_1, \ldots, Q_M)$ (see equation (2)). The coarse-grained entropy of $S$ is defined as

$$S_{cg}[\tau] = -\sum_{k=1}^{M} P(\Gamma_k, \tau) \ln P(\Gamma_k, \tau), \quad (45)$$

where $P(\Gamma_k, \tau)$ is the corresponding one-subsystem distribution function. This is the sum of partial entropies for each subsystem. The difference $S_{cg}[\tau] - S_{fg}[\tau]$ between the coarse-grained entropy (45) and fine-grained entropy (32) is non-negative (sub-additivity) and quantifies the relevance of correlations in $S$ [2,4].

For additionally motivating the definition (45), we can assume that the subsystems of $S$ were interacting for a finite time, and that $\tau$ is larger than this interaction time.

Note that the definition (45) is not the only possibility. There are (infinitely) many ways of doing coarse-graining, and thus many ways of defining non-equilibrium entropy\footnote{In particular, one can focus on certain macroscopic observables and define their physical, non-equilibrium entropy via maximization of information-theoretic entropy [4].}. The main advantage of (45) is that it allows one to see the entropy increase due to correlations (which is the main qualitative image behind the entropic formulation of the second law) [2,4]. To this end assume that initially the subsystems of $S$ are independent:

$$P(\Gamma, 0) = \prod_{k=1}^{M} P(\Gamma_k, 0). \quad (46)$$

This assumption specifies initial conditions needed for the existence of the thermodynamic arrow of time [2,4].

doi:10.1088/1742-5468/2008/04/P04001 14
If S starts from such a non-equilibrium state, and if the fine-grained entropy is constant in time due to the Liouville theorem, then one employs sub-additivity to get that the coarse-grained entropy is not decreasing in time:

\[ S_{cg}(t) \geq S_{fg}(t) = S_{fg}(0) = S_{cg}(0). \] (47)

Note that the rate of increasing for \( S_{cg}(t) \) was recently analyzed in [30].

However, once the Liouville theorem is not satisfied, \( S_{fg} \) can decrease in time and then (47) does not hold in general. There are other schemes for deriving the entropic formulation of the second law for different sets of initial states and for different definitions of the non-equilibrium entropy [4, 5, 7, 13]. All these derivations essentially use the Liouville theorem, so all of them do not apply to the present situation.

Note that there is a difference between inapplicability of the entropic formulation as compared to that of the Thomson formulation. Equation (47) shows that if the fine-grained entropy increases in time, the entropic formulation is satisfied. In contrast, the increasing fine-grained entropy does not yet ensure the validity of the Thomson formulation, as we discussed after (40).

7. The causal arrow

7.1. Reciprocity versus negligibility of the Born–Oppenheimer term

All the above anomalies with the second law are due to the fact that the Born–Oppenheimer term \( h(Q|Q_i, E_i) \) makes the dynamics of S not globally Hamiltonian. There are two related options for recovering this feature. First one can try to see whether the dependence of \( h(Q|Q_i, E_i) \) on \( Q_i \) can be neglected, \( h(Q|Q_i, E_i) \approx h(Q|E_i) \), but \( h(Q|E_i) \) still exerts a sizable force on S. Second, one can look for conditions where \( h(Q|Q_i, E_i) \) can be neglected as a whole. We shall now show that only the second option is consistent.

Employing (15), (16) as

\[ \Omega(E_i, Q_i) = \Omega(h(Q|E_i, Q_i), Q), \] (48)

and using (7) we get

\[ \partial_{Q_i} h(Q|E_i, Q_i) = \frac{\partial_{Q_i} \Omega(E_i, Q_i)}{\omega(h(Q|Q_i, E_i), Q)}, \] (49)

\[ \partial_{Q} h(Q|E_i, Q_i) = -\frac{\partial_{Q} \Omega(E, Q)|_{E=h(Q|E_i, Q_i)}}{\omega(h(Q|Q_i, E_i), Q)}. \] (50)

These equations show that there is a certain reciprocity—to be guessed already from (15), (16)—in the way \( h(Q|E_i, Q_i) \) depends on \( Q \) and \( Q_i \).

Let us demand that the Born–Oppenheimer term \( h(Q|E_i, Q_i) \) is independent from \( Q_i \). Since \( \omega(h(Q|Q_i, E_i), Q) \) is finite, this demand amounts to \( \partial_{Q_i} \Omega(E_i, Q_i) \rightarrow 0 \) for all \( E_i \) and \( Q_i \). This means requiring \( \partial_{Q} \Omega(E, Q)|_{E=h(Q|E_i, Q_i)} \rightarrow 0 \). Due to (50), this implies that \( h(Q|Q_i, E_i) \) reduces to a constant \( h(Q|Q_i, E_i) = E_i \), and—in addition—the energy of F does not change in time. We are thus led to assuming that there is no relevant interaction between S and F, a trivial option which is definitely not of interest to us.
We are thus left with the second option: for the timescales relevant for the dynamics of $S$ the Born–Oppenheimer Hamiltonian $h(Q_i, E_i)$ in (21) is negligible compared to the bare slow Hamiltonian $H_s(\Pi, Q)$. For this it is necessary to have

$$H_s(\Pi, Q) \gg h(Q_i, E_i).$$

In the absence of the Born–Oppenheimer term, the dynamics driven by $H_s$ is globally Hamiltonian, the Liouville theorem holds, and the second law is applicable to $S$; see the previous sections.

Using (15), (16) and (7) one calculates

$$\partial_{E_i} h(Q|E_i, Q_i) = \frac{\omega(E_i, Q_i)}{\omega(h(Q|E_i, Q_i), Q)} > 0.$$  (52)

This means that the Born–Oppenheimer term decreases with $E_i$. Since the RHS of (52) is normally $\sim \mathcal{O}(1)$, for satisfaction of (51) we have to require

$$H_s(\Pi, Q) \gg E_i.$$  (53)

We already saw this condition at the end of section 5.2 for a particular example. This example also shows that there may be situations where for sufficiently long times of the slow motion the Born–Oppenheimer force cannot be neglected, even though it is numerically small; see (31) in this context. In addition, there can be time limitations related to the validity of the timescale separation, and thus to the definition of the Born–Oppenheimer force; see the discussion after (30) in this context. Thus, at the moment we cannot give a fairly general estimate for the times for which the conditions (51) and (53) will be sufficient for neglecting the Born–Oppenheimer term.

### 7.2. The causal arrow

Equation (51) also means that the interaction between $S$ and $F$ gets the causal arrow: $S$ (cause) influences $F$ (effect), while $F$ does not influence $S$.

Thus we see that for the present system, the thermodynamic arrow and the causal arrow emerge simultaneously. Recall in this context the operational definitions of the causal arrow discussed in section 1.2.

### 8. Microcanonical ensemble and simplicity principle

After neglecting the Born–Oppenheimer term $h(Q_i, E_i)$ we recover a globally Hamiltonian behavior for the dynamics of $S$. In particular, the time average of the ergodic observables of $S$ can be described by a probability distribution:

$$P_s(\Gamma) = \frac{\delta(U_s - H_s(\Gamma))}{\int d\Gamma \delta(U_s - H_s(\Gamma))},$$  (54)

where $U_s$ is the slow energy. Since $S$ does not get a back-reaction from $F$, the energy $U_s$ is a constant determined by the initial conditions for the dynamics of $S$.

Recall that the very existence of (54) is related to neglecting the back-action of $F$ on $S$. For the same reason the probability distribution (54) is unconditional. The appearance
of (54) can be argued for following the lines of section 4. In this context we should assume that S with the Hamiltonian $H_s(\Pi, Q)$ is mixing and define the mixing time $\tau_s$ of S.

The distributions (17) and (54) can be combined into a microcanonical ensemble for describing the statistics of the overall system S + F for times larger than $\tau_s$, but smaller than the mixing time $\tau_{s+1}$ of the overall system:

$$P(\Gamma, z) = P_s(\Gamma)P_f(z|\Gamma).$$

(55)

It is understood that $Q_r$ needed in (17) for defining $P_f(z|\Gamma)$ is obtained (for given initial $\Gamma = (Q, \Pi)$) by solving the equations of motion (19) for S without the Born–Oppenheimer term.

Note that $P(\Gamma, z)$ in (55) can be obtained via sequential maximization of the conditional entropy $-\int dz\, P(z|\Gamma)\ln P(z|\Gamma)$ of F for fixed slow variables, and then maximization of the unconditional entropy $-\int d\Gamma\, P(\Gamma)\ln P(\Gamma)$ of S for fixed slow energy $U_s$. In this context it is not difficult to accept the idea that the microcanonical distribution is the simplest (least informative) one for a fixed value of energy.

On the other hand, the probability distributions $P(\Gamma|z)$ and $P(z)$—obtained from (55) via the Bayes formula—are not simple. They are not microcanonical, and in general they cannot even be obtained in a closed form.

Recalling that under condition (51) we identified S and F as the cause and effect, respectively, we get that the probability distributions $P(S)$ and $P(F|S)$ are simpler than $P(F)$ and $P(S|F)$. Let us see how this agrees with the notions of simplicity and plausibility proposed and studied in the modern machine-learning literature.

As proposed in [18], in causal reasoning one should tend to prefer the causal hypothesis $X_1 \rightarrow X_2$ ($X_1$ is the cause, and $X_2$ is its effect) if the factorization of $P(X_1, X_2)$ into $P(X_1)P(X_2|X_1)$ leads to significantly simpler terms $P(X_1)$ and $P(X_2|X_1)$ than the factorization into $P(X_2)P(X_1|X_2)$. Several causal inference principles based on different notions for the simplicity of conditional distributions have been proposed in the literature [18]–[20]. In [18] we have considered as the ‘most plausible’ ones those conditional distributions $P(X_2|X_1)$ which maximize the conditional entropy of $X_2$ given $X_1$ and are subject to the observed first and second moments $E(X_2), E(X_2^2), E(X_1X_2)$ (here $E(\cdot)$ denotes the expectation of a variable), after the marginal distribution $P(X_1)$ is given. In [19] we have considered $P(X_2|X_1)$ as simple (and ‘smooth’) if it maximizes the above conditional entropy subject to a function $f(X_1, X_2)$ whose norm $\|f\|$ in a so-called reproducing kernel Hilbert space (RKHS) [21] is small. Such RKHS norms are well known in machine learning. They provide a useful tool for giving preference to ‘smooth’ functions in various learning tasks, where smoothness is usually meant in the sense of small values of higher derivatives [21]. The inference principle studied in [20] is based on giving preference to conditionals of the form $P(X_2|X_1) = Q(\alpha X_2 - \beta X_1)$ with constants $\alpha, \beta$ and an arbitrary distribution $Q$. This form turns out to be a special case of the ‘most plausible’ conditional in [18] when applied to continuous variables. The general principle proposed in [18] also makes sense for discrete variables.

The common part of the rules in [18] and [19] is that one first maximizes the entropy of the cause variable subject to simple constraints, and then the conditional entropy of the effect, given the cause and subject to some other simple constraints. The fact that the order of the entropy maximization process is given by the causal order coincides with what we observe in the present paper: the microcanonical distributions $P(S)$ and $P(F|S)$
Relating the thermodynamic arrow of time to the causal arrow

are also given by first maximizing the entropy of the cause and then maximizing the conditional entropy of the effect, given the cause and subject to simple constraints. Here, the constraints are ‘simple’ in the sense that they consist only of a fixed energy value for S and an S-dependent energy value for F.

The causal arrow persists in the global microcanonical equilibrium which—if the overall system S + F is mixing with a time \( \tau_{s+f} \)—is established for \( t \gg \tau_{s+f} \):

\[
P_{eq}(\Gamma, z) = \frac{\delta(E - H_s(\Gamma) - H(Q, z))}{\int d\Gamma' dz' \delta(E - H_s(\Gamma) - H(Q, z))},
\]

where \( E \) is the total energy. Equation (56) is a stationary distribution. The no-back-action condition (51) is now substituted by its equilibrium analog

\[
H_s(\Pi, Q) \gg H(Q, z).
\]

However, once the slow Hamiltonian is much larger than the fast Hamiltonian, we expect that the partial probability \( P_{eq}(\Pi, Q) \) will be close to \( P_s(\Pi, Q) \) in (54). Indeed, once \( H(Q, z) \) is small, the overall energy \( E \) in (56) should be nearly canceled by the bare slow Hamiltonian \( H_s(\Pi, Q) \), so that \( P_{eq}(\Pi, Q) \) is proportional to a smeared delta function concentrated at \( E = H_s(\Pi, Q) \). For calculating observables (at small \( H(Q, z) \)) this is the same as \( P_{eq}(\Pi, Q) \propto \delta(E - H_s(\Pi, Q)) \).

As for the conditional probability \( P_{eq}(z|\Pi, Q) = P_{eq}(z|\Gamma) \), it can always be written as

\[
P_{eq}(z|\Gamma) = \frac{\delta(E - H_s(\Gamma) - H(Q, z))}{\int dz' \delta(E - H_s(\Gamma) - H(Q, z))}.
\]

Here \( E - H_s(\Gamma) \) is, of course, not the Born–Oppenheimer energy \( h(Q_{\tau}|E_i, Q) \) that shows up in the non-equilibrium distribution (17).

9. Discussion

9.1. Summary

We studied a Hamiltonian system that consists of a slow subsystem S and a fast subsystem F; see section 2. The separation into slow versus fast is one of the basic ways of defining autonomous systems in the natural sciences [22]. In particular, the effective dynamics of slow subsystems is studied in a great variety of different fields: atomic and molecular physics, semi-classical physics (including semi-classical gravity), physical chemistry, synergetics, economics, etc.

Our main purpose was relating two seemingly different issues: (i) the causal arrow—or unidirectional influence—where S influences F, but does not get a back-action; (ii) the thermodynamic arrow of time (second law) for the system. Since the applicability of the second law to F is well known [8,11], we focused on the second law as applied to the autonomous, energy conserving, Hamiltonian dynamics of S. The presence of F is reflected in the dynamics of S via an additional Born–Oppenheimer term in the Hamiltonian of S. This term emerged during the tracing out of F, and it depends on the initial coordinate of S; see section 5. Thus, different initial coordinates of S have different Hamiltonians: the dynamics of S is not globally Hamiltonian. The cause of this is that due to the timescale separation the dynamics of F does have an adiabatic invariant (effective conservation law); see section 3.

doi:10.1088/1742-5468/2008/04/P04001 18
The specific features of the Born–Oppenheimer term make the basic formulations of the second law inapplicable to the dynamics of S. These statements of the second law are (i) the Thomson formulation, which states that no work can be extracted by means of a cyclic Hamiltonian process (driven by an external source of work) if the initial conditions of S are thermal and (ii) entropic formulation, which claims that the coarse-grained entropy of S does not decrease provided that S starts from a low entropy state. There are two mechanisms for this inapplicability. First, the Liouville theorem (i.e., conservation of the fine-grained entropy) does not hold for a non-globally Hamiltonian dynamics: the fine-grained entropy can both increase and decrease over the course of time. The second mechanism is efficient for the Thomson formulation only and has to do with the behavior of the Born–Oppenheimer term under a cyclic Hamiltonian driving; see section 6 for details.

As we argued in section 7.1, the Born–Oppenheimer term has a certain reciprocity feature. Its basic implication for our purposes is that the only way to recover a globally Hamiltonian dynamics for S is to neglect the Born–Oppenheimer term as compared to the bare Hamiltonian of S. In this way we neglect the influence of F on S, but, importantly, the influence of S on F is not neglected and can be sizable. Once the Born–Oppenheimer term can be neglected, the basic formulations of the second law naturally apply to S. Thus we see that the emergence of the thermodynamic arrow (second law) for S is closely related to the causal arrow: S acts on F, but does not get back-action.

Finally, in section 8, we studied our results in the context of a causal inference principle proposed recently in machine learning [18]. This principle plausibly infers the cause–effect relation between two stochastic variables, and it intends to cover especially those situations where more standard causal inference procedures do not apply. If we assume that S and F are mixing systems, under the causal arrow condition they are described by the microcanonical probability distribution $P(S)$ and the conditional microcanonical distribution $P(S|F)$. Now the factorization of the joint probability $P(cause = S, effect = F)$ into $P(cause)P(effect|cause)$ leads here to simpler expressions than the factorization into $P(effect)P(cause|effect)$. This is the core of the inference principle proposed in [18], and we conclude that this principle is validated within the model studied here.

9.2. Post-adiabatic corrections and Brownian motion

Above we discussed the implications of the non-globally Hamiltonian feature of the Born–Oppenheimer potential. In the next-to-leading order of the adiabatic approach there are two additional forces generated by the fast system F on the slow system S. Those are the deterministic friction\footnote{Friction is a non-Hamiltonian force that depends on the velocity of the particle, e.g., it appears in the RHS of (19) as $-\gamma \dot{Q}$: $(d/d\tau)\Pi = -\partial_{\xi}\mathcal{H}_\xi - \gamma \dot{Q}$, where $\gamma > 0$ is the coefficient of friction, which depends of course on the coupling between S and F. (Equation (20) does not change.) The friction force (without the noise) leads to energy dissipation: $(d/d\tau)\mathcal{H}_\xi(\Pi(t),Q(t)) = -\gamma \dot{Q}^2 < 0$.} and the stochastic force (noise), which is related to the friction by an analog of the fluctuation-dissipation relation [31,32]\footnote{In the same order there is yet another, non-potential force which is analogous to the magnetic field (i.e., it does not lead to energy dissipation) [31]. This will however not play any important role in the subsequent discussion.}. These forces arise already for a finite, but chaotic F, and in the adiabatic limit they have an additional smallness as compared to the Born–Oppenheimer term; the latter scales as $O(1)$, while the friction
Relating the thermodynamic arrow of time to the causal arrow

and the (contribution of the) stochastic force scale as $O(\tau_f/\tau_s)$, where $\tau_f$ and $\tau_s$ are, respectively, the fast and slow characteristic times.

The appearance of the above two additional forces is to be expected by the analogy to the physics of Brownian motion, where a slow Brownian particle interacts with fast particles of a macroscopic thermal bath [33]. The bath exerts on the Brownian particle the same forces of friction and noise, while the Born–Oppenheimer force is normally neglected in the standard theories of Brownian motion [33]. Those standard theories are known to satisfy the second law at least when the dynamics of the Brownian particle is Markovian\(^\text{14}\) (i.e., the noise is white) [7,33]. By this we mean, in particular, the statement on the monotonic decay of free energy with time [7,33] (while for an energy conserving dynamics the second law is frequently expressed via entropy, for an open system interacting with an equilibrium thermal bath it is more customary to use the free energy).

The same conclusion on the obeying of the second law was reached for the forces generated by a small system $F$, again without taking into account the non-globally Hamiltonian feature of the Born–Oppenheimer force [31,32]. It is clear that these conclusions are consistent with the results obtained in the present work, though we do not consider the dissipative regime, i.e., we restrict ourselves to the proper adiabatic regime (and not very long times), where the influence of the two post-adiabatic forces can be neglected. What happens when the above two dissipative forces are studied together with the Born–Oppenheimer term (taking seriously its non-globally Hamiltonian feature) is in our opinion an open problem. The same question can be asked for the standard theories of Brownian motion, where the Born–Oppenheimer term is also normally neglected (sometimes simply by design; see, e.g., [34]).

Acknowledgment

This work was supported by Volkswagenstiftung grant ‘Quantum Thermodynamics: Energy and information flow at nanoscale’.

References

[1] Reichenbach H, 1956 The Direction of Time (Berkeley: California University Press)
[2] Penrose O and Percival I C, 1962 Proc. Phys. Soc. 79 605
[3] Schulman L S, 1997 Times Arrows and Quantum Measurement (Cambridge: Cambridge University Press)
[4] Balian R, 1992 From Microphysics to Macrophysics vol I and II (Berlin: Springer)
[5] Jaynes E T, 1965 Am. J. Phys. 33 391
[6] Zeh H D, 2001 The Physical Basis of the Direction of Time (Berlin: Springer)
[7] Lindblad G, 1983 Non-Equilibrium Entropy and Irreversibility (Dordrecht: Reidel)
[8] Bassett I M, 1978 Phys. Rev. A 18 2356
[9] Lenard A, 1978 J. Stat. Phys. 19 575
[10] Janzing D, Wocjan P, Zeier R, Geiss R and Beth Th, 2000 Int. J. Theor. Phys. 39 2217
[11] Allahverdyan A E and Nieuwenhuizen Th M, 2002 Phys. Rev. B 66 115309
[12] Allahverdyan A E and Nieuwenhuizen Th M, 2007 Phys. Rev. E 75 051124
[13] Allahverdyan A E and Nieuwenhuizen Th M, 2005 Phys. Rev. E 71 046107
[14] The entropic analog of the minimum work principle was studied in Campisi M, 2007 Preprint 0704.2567 [cond-mat.stat-mech]
[15] Campisi M, 2005 Stud. Hist. Phil. M. P. 36 275
[16] Allahverdyan A E and Nieuwenhuizen Th M, 2005 Preprint cond-mat/0508089

\(^\text{14}\) See [13] for a more general perspective on the relevance of the Markovian feature for the second law.
Relating the thermodynamic arrow of time to the causal arrow

[14] Gemmer J, Michel M and Mahler G, 2004 Quantum Thermodynamics—Emergence of Thermodynamic Behavior within Composite Quantum Systems (Springer Lecture Notes in Physics vol 657) (Berlin: Springer)

[15] Penrose R, 1994 J. Stat. Phys. 77 217
Wald R M, 2005 Preprint gr-qc/0507094
Carroll S M and Chen J, 2004 Preprint hep-th/0410270
Carroll S M and Chen J, 2005 Preprint gr-qc/0505037

[16] Hoffding H, 1955 History of Modern Philosophy (New York: Dover)

[17] Pearl J, 2000 Causality (Oxford: Oxford University Press)

[18] Sun X, Janzing D and Schölkopf B, Causal inference by choosing graphs with most plausible Markov Kernels, 2006 Proc. 9th Int. Symp. on Art. Intell. and Math. (Florida: Fort Lauderdale) pp 1–10 http://anytime.cs.umass.edu/aimath06/proceedings/P16.pdf

[19] Sun X, Janzing D and Schölkopf B, Causal reasoning by evaluating the complexity of conditional densities with kernel methods, 2007 Neurocomputing 71 1248

[20] Kano Y and Shimizu S, Causal inference using non-normality, 2003 ISM Report on Research and Education No 17 (Tokyo: The Institute of Statistical Mathematics) pp 261–70

[21] Schölkopf B and Smola A, 2002 Learning with Kernels (Cambridge, MA: MIT Press)

[22] Haken H, 1983 Synergetics, an Introduction: Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry, and Biology (Berlin: Springer)

[23] van Kampen N G, 1971 Physica 53 98

[24] Hertz P, 1910 Ann. Phys., Lpz. 33 225
Hertz P, 1910 Ann. Phys., Lpz. 33 537
Kasuga T, 1961 Proc. Japan. Acad. 37 366

[25] Munster A, 1969 Statistical Thermodynamics vol 1 (Berlin: Springer)

[26] Rugh H H, 2001 Phys. Rev. E 64 055101

[27] Ott E, 1979 Phys. Rev. Lett. 42 1628
Jarzynski C, 1993 Phys. Rev. Lett. 71 839

[28] Sagdeev R Z, Usikov D A and Zaslavsky G M, 1988 Nonlinear Physics (Philadelphia: Harwood)

[29] Lichtenberg A J and Lieberman M A, 1991 Regular and Chaotic Dynamics (New York: Springer)

[30] Falcioni M, Palatella L, Pigolotti S, Rondoni L and Vulpiani A, 2007 Physica A 385 170

[31] Berry M V and Robbins J M, 1993 Proc. R. Soc. A 442 659

[32] Jarzynski C, 1995 Phys. Rev. Lett. 74 2937

[33] Risken H, 1984 The Fokker–Planck Equation (Berlin: Springer)

[34] Weiss U, 1993 Quantum Dissipative Systems (Singapore: World Scientific)