

# How do macrostates come about?

Márton Gömöri\*, Balázs Gyenis†, and Gábor Hofer-Szabó‡

## Abstract

This paper is a further consideration of Hemmo and Shenker’s (2012) ideas about the proper conceptual characterization of macrostates in statistical mechanics. We provide two formulations of how macrostates come about as elements of certain partitions of the system’s phase space imposed on by the interaction between the system and an observer, and we show that these two formulations are mathematically equivalent. We also reflect on conceptual issues regarding the relationship of macrostates to distinguishability, thermodynamic regularity, observer dependence, and the general phenomenon of measurement.

**Keywords:** macrostates, distinguishability, thermodynamic regularity

## 1 Introduction

Macrostates are distinguished subsets of a system’s phase space. They play an essential role in statistical mechanics since they are identified—at least in the Boltzmann program—with thermodynamic states and thus provide the basis for the statistical mechanical explanation for thermodynamic phenomena. It is a crucial task for statistical mechanics to give a physical explanation for why certain sets of microstates become distinguished, or in short, how macrostates come about.

As an answer to this question, in their illuminating book *The Road to Maxwell’s Demon, Conceptual Foundations of Statistical Mechanics* Meir Hemmo and Orly Shenker discern two characteristic features of macrostates:

1. Microstates in a macrostate are indistinguishable, while macrostates are distinguishable to a human observer.

“One kind of physical property according to which sets of microstates can be defined is distinguishability by a given observer: in general, observers are unable to distinguish between individual microstates, but can distinguish between certain sets of microstates; and each distinguishable set of indistinguishable microstates forms a macrostate.” (Hemmo and Shenker, 2012, p. 95)

---

\*Research Center for the Humanities, Budapest, email: gomori.marton@btk.mta.hu

†Research Center for the Humanities, Budapest, email: gyenis.balazs@btk.mta.hu

‡Research Center for the Humanities, Budapest, email: szabo.gabor@btk.mta.hu

2. Macrostates feature in thermodynamic regularities (such as the ideal gas law).

“Another kind of physical property of interest, shared by microstates, is the one that gives rise to the thermodynamic regularities. Certain sets of microstates exhibit this particular kind of regularity: all the microstates in these sets appear to satisfy the same laws, described by the theory of thermodynamics.” (Hemmo and Shenker, 2012, p. 95)

These two characterizations of the macrostates are conceptually distinct: the first characterization refers to the system’s relation to an external observer while the second one refers only to features inherent to the system. Still, there is a significant coincidence between these two features of the macrostates:

“It is a contingent fact about the structure of human beings as observers, that there is a useful degree of overlap between the sets that satisfy the regularities and the sets that correspond to our observation capabilities.” (Hemmo and Shenker, 2012, p. 96)

Hemmo and Shenker keep emphasizing that despite being observer-relative macrostates are objective. Their book provides an insightful analysis of what distinguishability means in terms of the physical interaction of the observer and the target system. The key idea of their analysis is that the interaction of the system and the observer brings about a one-to-one correlation between certain sets of microstates in the system’s phase space and the observer’s phase space. It is these sets of the system’s phase space that they call the *macrostates* of the system relative to the observer.

The main aim of the present paper is to provide a precise mathematical description of how macrostates come about via the correlation of states of an observer and a system. We hope that this analysis will contribute to making Hemmo and Shenker’s illuminating ideas a bit sharper.

The paper is structured as follows. In Section 2 we give an intuitive introduction to the idea behind the mechanism of the coming about of macrostates. In Section 3 we provide a mathematically rigorous analysis of the same mechanism together with a mathematically equivalent characterization of macrostates. In Section 4 we provide a formal description of the observer-relativeness of macrostates. Readers not interested in technicalities can go directly to Section 5 where the results of the previous two sections will be summarized in an informal way. In Section 5 we reflect on conceptual issues regarding the relationship of macrostates characterized in the above way to distinguishability, thermodynamic regularity and the general phenomenon of measurement. We conclude in Section 6.

## 2 Macrostates – triggering intuition

Let  $O$  be an observer and  $S$  a physical system. Let  $X_O$  and  $X_S$  denote the *phase space* of  $O$  and  $S$ , respectively. Generally we will not assume that  $X_O$  and  $X_S$  have any mathematical

structure. They will simply be sets if not explicitly stated otherwise. Let  $O + S$  denote the *joint system* of the observer and the system, and let  $X_{O+S}$  denote the phase space of the joint system. We stipulate that a *microstate*  $x \in X_{O+S}$  of the joint system is given by a pair  $(x_O, x_S)$  where  $x_O \in X_O$  is a *microstate* of the observer and  $x_S \in X_S$  is a *microstate* of the system; that is  $X_{O+S} = X_O \times X_S$  is a Cartesian product. Sometimes we will refer to  $x_O$  and  $x_S$  as the *projection* of  $x$  onto  $X_O$  and  $X_S$ , respectively. Similarly, for any *microregion*  $B \subseteq X_{O+S}$  let  $B_O$  and  $B_S$  denote the projection of  $B$  onto  $X_O$  and  $X_S$ , respectively.

The possible microstates of the joint system  $O + S$  are often confined to a subregion of  $X_{O+S}$  due to various physical conditions. Denote this *accessible region* of microstates of the joint system by  $A$  and its projections onto  $X_O$  and  $X_S$  by  $A_O$  and  $A_S$ , respectively. The exact “shape” of the accessible region  $A$  depends on the nature of the physical interaction between the observer and the system and, as we will shortly see, it is playing a crucial role in the coming about of macrostates.

How do macrostates come about?

The central idea, due to Hemmo and Shenker (2012, Ch. 5), is that (i) macrostates of a system  $S$  are relative to an observer  $O$ ; (ii) they emerge from the many-to-many type correlation between the microstates of  $O$  and  $S$ ; and, most importantly, (iii) this correlation is established by the accessible region  $A$  of the joint system. In the next two sections we will give a mathematically precise formulation of these ideas; here we just trigger intuition.

Consider Fig. 1, borrowed from the book of Hemmo and Shenker.

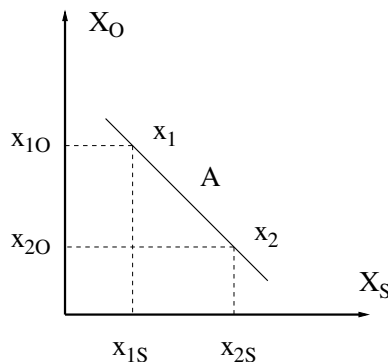


Figure 1: One-to-one correlation between the microstates of  $O$  and  $S$  in a connected accessible region  $A$ .

Here the microstates of  $O$  are depicted along the vertical axis of the diagram and the microstates of  $S$  along the horizontal axis. The accessible region  $A$  of the joint system is a slant line. Note that the axes and the straight line are just for illustration since the phase spaces have no linear, metric or topological structure.

Now, the accessible region  $A$  of the joint system is such that it establishes a one-to-one correlation between the microstates of  $O$ . In other words, the outer physical conditions

and the nature of the interaction between the observer and the system are such that if the observer  $O$  is in a certain microstate, the system  $S$  is forced to be in one particular microstate.

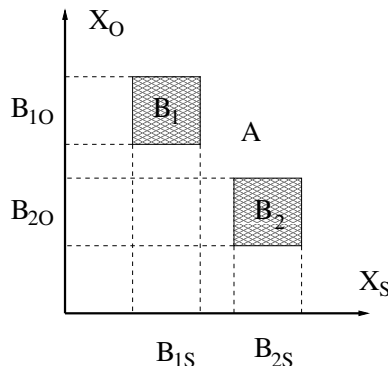


Figure 2: Many-to-many correlation between the microstates of  $O$  and  $S$  in a disconnected accessible region  $A$ .

Consider now another situation where the accessible region  $A$  of the joint system consists of two disconnected regions,  $B_1$  and  $B_2$ . (See Fig. 2.) What kind of constraints arise due to  $A$  between the microstates of  $O$  and  $S$ ? Obviously, if the microstate of  $O$  is in  $B_{1O}$ , then the microstate of  $S$  cannot be in  $B_{2S}$ . Thus, the microstates of the two subsystems cluster into two groups: microstates of the system in  $B_{1S}$  correlate with microstates of the observer in  $B_{1O}$ , and microstates in  $B_{2S}$  correlate with microstates in  $B_{2O}$ .

Observe that the difference between the two cases lies completely in the difference between the two accessible regions. In the first case the accessible region maps the microstates of the observer to the microstates of system in a one-to-one way, whereas in the second case the accessible region maps only one particular partition of the phase space of the observer, namely  $\{B_{1O}, B_{2O}\}$  to one particular partition of the phase space of the system, namely  $\{B_{1S}, B_{2S}\}$ . Stating it differently, in the first case the accessible region maps *any* partition of the phase space of the observer to the corresponding partition of the phase space of the system; whereas in the second case the accessible region maps only *specific* partitions of the phase space of the observer to *specific* partitions of the phase space of the system in a one-to-one manner.

And this is the point where macrostates enter the scene. *Macrostates* are elements of specific partitions of  $X_S$ , namely of those partitions which are mapped by the accessible region of the joint system  $O + S$  into partitions of  $X_O$  in a one-to-one way.

What is the idea behind this definition of macrostates?

Suppose you measure the temperature of a container of gas. Measuring temperature means that one introduces an interaction between a measuring device, the thermometer, and the system. The nature of this interaction fixes which joint microstates the joint system can have. The different mercury levels partition the phase space of the thermometer

regarded as a system of particles. But in order for a thermometer to count as a *measuring device* and for the mercury levels to count as *measurement outcomes* it is also needed that the different levels of mercury tell something about the state of the measured system. A given mercury level is compatible with many microstates of the thermometer; and it will not specify the microstate of the system either. However, what is required in the above definition is that the level of mercury should indicate at least in which *set* of microstates the system is. To do this, the sets pertaining to different mercury levels cannot overlap—otherwise the thermometer would not serve as a good measuring device; and they together should cover the phase space of the system—otherwise some microstates of the system could not be indicated.

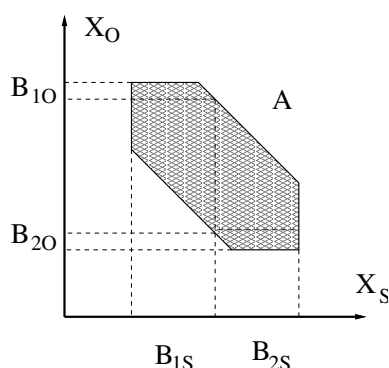


Figure 3: Many-to-many correlation with disjoint sets on the phase space of  $O$ .

One could, however, also take a more liberal stance towards macrostates. Suppose that the phase space of the observer cannot be partitioned entirely but there is a *set of disjoint subsets* such that their image *via* the the accessible region in the above sense forms a partition of the phase space of the system. (See Fig. 3.) The difference between this case and the previous one is that here the sets of those microstates of the observer which stand in a one-to-one relation with a certain partition of the system's phase space do not form a partition. Some microstates of the observer do not indicate any macrostate of the system.

An even more liberal definition of the macrostates would be to relax the demand of being a partition both for the phase space of the observer and also for the system and only to demand that disjoint subsets of  $O$  be mapped into disjoint subsets of  $S$  *via* the accessible region. (See Fig. 4.)

Finally, one could demand even less. For simplicity let  $f$  and  $g$  be real-valued functions on  $X_O$  and  $X_S$ , respectively, and map out the graph composed of the pair of values  $(o, s)$  where  $o = f(x_O)$ ,  $s = g(x_S)$ , and  $x$  runs through all points in the accessible region  $A$ . Suppose that the resulting graph has a shape which allows statistical inference from (sets of)  $o$ -values to (sets of)  $s$ -values. To the extent this statistical inference is reliable it is possible to associate (sets of) microstates of the observer with (sets of) microstates of the system, and thus in this weak sense we could speak of the observer being able to discern

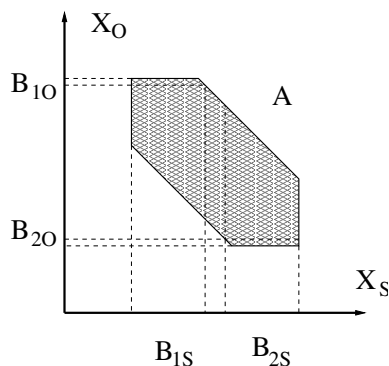


Figure 4: Many-to-many correlation with disjoint sets on the phase spaces of  $O$  and  $S$ .

different macrostates. In case the graph is a graph of a one-to-one function we get back our first, strict definition of a macrostate, but it is easy to see that there could be other physical situations (such as the one depicted on Fig. 5) which could still warrant relying on this weaker concept.

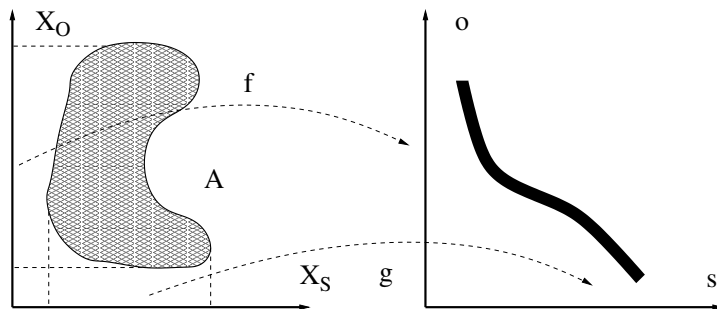


Figure 5: Inverse functions  $f^{-1}$  and  $g^{-1}$  “almost” partition the phase spaces  $X_O$  and  $X_S$ .

In general, if one follows the path charted out by Hemmo and Shenker, the macrostate concept needs to be tailored to physical characteristics of the measurement procedure. In cases when the joint system needs to be in an equilibrium microstate in order for the measurement to count as such—in fact this is the case in the temperature measurement example above—, one would either need to apply one of the weaker notions of a macrostate, or apply the partition-to-partition macrostate notion but impose further restrictions on the accessible region  $A$  in order to exclude non-equilibrium microstates. The weakest macrostate notion may be warranted when, with an even more emphasized epistemic bent, one takes into account measurement errors, noises, environmental effects, fuzziness of macroscopic concepts, and so on.

In what follows we focus on the notion of macrostate only in the first, strongest sense, namely when *partitions* are transformed into one another in a one-to-one way. In the following two sections we will make the notion of macrostate mathematically precise and return to their physical interpretation in Section 5.

### 3 Macrostates – the formal definition

Consider a joint system  $O + S$  with accessible region  $A$ . Denote by  $2^A, 2^{A_O}$  and  $2^{A_S}$  the *power set* of the accessible region  $A$ , and the *power set* of its projections  $A_O$  and  $A_S$ , respectively. The accessible region  $A$  will uniquely define the following maps:

$$\mathcal{A}^o : A_O \rightarrow 2^A; \quad \mathcal{A}^o(x_O) := \{x' \in A \mid x'_O = x_O\} \quad (1)$$

$$\mathcal{A}^s : A_S \rightarrow 2^A; \quad \mathcal{A}^s(x_S) := \{x' \in A \mid x'_S = x_S\} \quad (2)$$

$$\mathcal{A}^{os} : A_O \rightarrow 2^{A_S}; \quad \mathcal{A}^{os}(x_O) := (\mathcal{A}^o(x_O))_S \quad (3)$$

$$\mathcal{A}^{so} : A_S \rightarrow 2^{A_O}; \quad \mathcal{A}^{so}(x_S) := (\mathcal{A}^s(x_S))_O \quad (4)$$

Intuitively,  $\mathcal{A}^o(x_O)$  is picking out those microstates of the joint system which are in  $A$  and the projection of which onto  $X_O$  is  $x_O$ .  $\mathcal{A}^{os}$  is simply the composition of  $\mathcal{A}^o$  and the projection onto  $X_S$ ; it maps elements of  $A_O$  onto subsets of  $A_S$ . The functions  $\mathcal{A}^s$  and  $\mathcal{A}^{so}$  are defined similarly.

One can also “lift up” the above maps to the level of power sets:

$$\mathcal{A}^o[\ ] : 2^{A_O} \rightarrow 2^A; \quad \mathcal{A}^o[B_O] := \{x' \in A \mid x'_O \in B_O\} \quad (5)$$

$$\mathcal{A}^s[\ ] : 2^{A_S} \rightarrow 2^A; \quad \mathcal{A}^s[B_S] := \{x' \in A \mid x'_S \in B_S\} \quad (6)$$

$$\mathcal{A}^{os}[\ ] : 2^{A_O} \rightarrow 2^{A_S}; \quad \mathcal{A}^{os}[B_O] := (\mathcal{A}^o[B_O])_S \quad (7)$$

$$\mathcal{A}^{so}[\ ] : 2^{A_S} \rightarrow 2^{A_O}; \quad \mathcal{A}^{so}[B_S] := (\mathcal{A}^s[B_S])_O \quad (8)$$

Let  $\mathbf{P}^A, \mathbf{P}^{A_O}$  and  $\mathbf{P}^{A_S}$  denote the *set of partitions* of  $A, A_O$  and  $A_S$ , respectively. Let  $P^{A_O} \in \mathbf{P}^{A_O}$  and  $P^{A_S} \in \mathbf{P}^{A_S}$  be two partitions. Note that  $\mathbf{P}^{A_O} \neq \mathbf{P}_O^A$ , that is the partitions of  $A_O$  are *not* the projections of the partitions of  $A$ . We will apply the maps (5)-(8) also to the partitions:

$$\mathcal{A}^{os}[P^{A_O}] := \{\mathcal{A}^{os}[B_O] \mid B_O \in P^{A_O}\}$$

Note that  $\mathcal{A}^{os}[P^{A_O}]$  is not necessarily a partition of  $A_S$ .

With these notations in hand we can now define the notion of macrostates introduced in the previous section:

**Definition 1.** Let  $\{P^{A_O}, P^{A_S}\}$  be a pair of partitions such that  $\mathcal{A}^{os}[P^{A_O}] = P^{A_S}$  and  $\mathcal{A}^{so}[P^{A_S}] = P^{A_O}$ . Then we call an element of  $P^{A_S}$  a *macrostate* of the system  $S$  relative to  $O$  with accessible region  $A$ .

Next we provide a characterizations of macrostates which is equivalent to the above definition. To this aim we introduce the notion of projective connectedness.

**Definition 2.** Let  $x$  and  $x'$  be microstates in  $A$ . We call  $x$  and  $x'$  *projectively connectible* in  $A$ ,  $x \sim x'$ , if there exists a finite sequence  $\{x_n\}_{n=1}^N$  with all  $x_n \in A$  such that  $x_1 = x$ ,  $x_N = x'$  and for any  $x_n$  either  $(x_n)_O = (x_{n+1})_O$  or  $(x_n)_S = (x_{n+1})_S$ . In other words,  $x \sim x'$

iff  $x$  and  $x'$  can be connected by moving within  $A$  only along projections onto  $X_O$  and  $X_S$ , respectively. A region  $B$  in  $A$  is called *projectively connectible* if  $x \sim x'$  for any  $x, x' \in B$ . (See Fig. 6.) A set  $\{B_i\}$  of regions in  $A$  is called *mutually projectively unconnectible* if  $x_i \not\sim x_j$  for any  $x_i \in B_i$  and  $x_j \in B_j$  with  $i \neq j$ .<sup>1</sup>

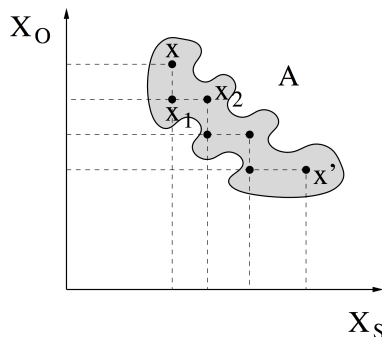


Figure 6: A projectively connectible region.

Denote a partition of  $A$  by  $P^{A\sim}$  if its elements are mutually projectively unconnectible. Denote the set of such partitions of  $A$  by  $\mathbf{P}^{A\sim}$ . Note that  $P_{max}^{A\sim}$ , the partition generated by the equivalence classes of the equivalence relation  $\sim$  is only one (the finest) partition of all partitions  $P^{A\sim}$ . A partition  $P$  is called *finer* than a partition  $P'$  if for any  $B' \in P'$  there exists a set  $\{B_i\}$  of elements in  $P$  such that  $B' = \cup_i B_i$ .  $P$  is *strictly finer* than  $P'$  if at least one set  $\{B_i\}$  contains at least two elements.

Let  $P_O^{A\sim}$  and  $P_S^{A\sim}$  denote the projections of  $P^{A\sim}$  onto  $X_O$  and  $X_S$ , respectively. Now, we formulate a new definition of macrostates.

**Definition 3.** Let  $P^{A\sim}$  be a mutually projectively unconnectible partition. Then we call an element of  $P_S^{A\sim}$  an *unconnectibility macrostate* of the system  $S$  relative to  $O$  with accessible region  $A$ .

However, as the following Proposition shows, Definitions 1 and 3 of macrostates are equivalent. Therefore we omit the adjective “unconnectibility.”

**Proposition 1.** A subset of  $A_S$  is a macrostate iff it is an unconnectibility macrostate.

*Proof.* For the proof of Proposition 1 see the Appendix.

---

1. Our definition of projectively connectible points allow only finite number of steps of moving along projections. In case the phase space has a topological structure the definition could be straightforwardly extended to allow for the points to be projectively connected in the infinite limit. Such extension of the definition would not alter the result of Proposition 1 and Proposition 2.



## 4 Macrostates are observer-relative

Macrostates are observer-relative. If instead of  $O$  another observer,  $O'$ , interacts with the system then the partition of the system's phase space induced by the interaction can be different. Measuring a container of gas by a thermometer and by a manometer do not partition the phase space of the gas in a same way.

Observer-dependence can be understood in two different ways. Let  $O$  and  $O'$  be two different observers who *separately* perform measurements on the same system  $S$ . Let  $X_{O+S}$  and  $X_{O'+S}$  denote the phase space of the joint system  $O + S$  and  $O' + S$ , respectively. Now, due to the different nature of interactions between the system and the different observers the accessible region  $A_{O+S}$  of the joint system  $O + S$  can be different from the accessible region  $A_{O'+S}$  of the joint system  $O' + S$ . Since macrostates of the system  $S$  are generated by the accessible region of the joint system, there is no a priori guarantee that the macrostates of  $S$  relative to  $O + S$  and  $O' + S$  will be the same.

Observer-dependence can also be understood, however, in a different way. Let  $O + O' + S$  be the joint system of the *two* observers who perform a *joint* measurement on the system  $S$ . This is the case when we perform the temperature measurement and the pressure measurement simultaneously. Let  $X_{O+O'+S}$  denote the phase space and let  $A$  denote the accessible region of the joint system. Let the phase spaces and the microstates of the subsystems be denoted just as above.

What are the macrostates of the triply joint system  $O + O' + S$ ? One can proceed here either in a permissive or in a restrictive way. The permissive characterization of the macrostates is to say that there is a partition  $P^{A_O}$  of  $A_O$ , a second partition  $P^{A_{O'}}$  of  $A_{O'}$  and *two more* partitions  $P^{A_S}$  and  $P'^{A_S}$  of  $A_S$  such that the (common) accessible region  $A$  of the joint system  $O + O' + S$  sends the partition  $P^{A_O}$  into  $P^{A_S}$ , and the partition  $P^{A_{O'}}$  into  $P'^{A_S}$  in a one-to-one way. Hence we obtain the following definition:

**Definition 4.** Let  $\{P^{A_O}, P^{A_S}\}$  and  $\{P^{A_{O'}}, P'^{A_S}\}$  be two pairs of partitions of  $A_O$ ,  $A_{O'}$  and  $A_S$ , respectively, such that

$$\mathcal{A}^{os}[P^{A_O}] = P^{A_S} \quad \mathcal{A}^{so}[P^{A_S}] = P^{A_O} \quad (9)$$

$$\mathcal{A}^{o's}[P^{A_{O'}}] = P'^{A_S} \quad \mathcal{A}^{so'}[P'^{A_S}] = P^{A_{O'}} \quad (10)$$

Then we call an element of  $P^{A_S}$  and  $P'^{A_S}$  a *macrostate* of the system  $S$  relative to  $O$  and  $O'$ , respectively, within the joint system  $O + O' + S$  with accessible region  $A$ .

The restrictive characterization would be, however, to demand that both observers generate the *same partition* of  $A_S$ , that is to demand that there is only *one common set of macrostates* of  $S$  associated to both observers:

**Definition 5.** Let  $\{P^{A_O}, P^{A_{O'}}, P^{A_S}\}$  be a triple of partitions of  $A_O$ ,  $A_{O'}$  and  $A_S$ , respectively, such that

$$\mathcal{A}^{os}[P^{A_O}] = P^{A_S} \quad \mathcal{A}^{so}[P^{A_S}] = P^{A_O} \quad (11)$$

$$\mathcal{A}^{o's}[P^{A_{O'}}] = P^{A_S} \quad \mathcal{A}^{so'}[P^{A_S}] = P^{A_{O'}} \quad (12)$$

Then we call an element of  $P^{As}$  a *common macrostate* of the system  $S$  relative to both  $O$  and  $O'$  within the joint system  $O + O' + S$  with accessible region  $A$ .

Obviously, whether two observers generate one common partition, two different partitions, or no partition at all essentially depends on the accessible region  $A$  of the joint system  $O + O' + S$ . In what follows we give examples for all three cases:

**Example 1.** A joint system  $O + O' + S$  with no (non-trivial) macrostates.

Let  $A$  be such that for any  $x, y \in A$

- (i)  $\mathcal{A}^{so}[X_S] = X_O$ ;
- (ii)  $\mathcal{A}^{so}(x_S) \neq \mathcal{A}^{so}(y_S)$  if  $x_S \neq y_S$ ;
- (iii)  $\mathcal{A}^{s'o'}[\{x_S\}] = X_{O'}$ .

(See Fig. 7.)

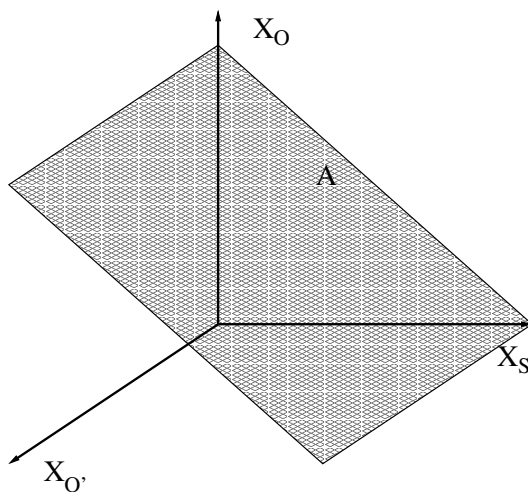


Figure 7: An accessible region with no (non-trivial) partitions  $\{P^{A_{O'}}, P^{A_S}\}$ .

In this case there are no (non-trivial) partitions  $\{P^{A_{O'}}, P^{A_S}\}$  which would satisfy (10).

**Example 2.** A joint system  $O + O' + S$  with separate partitions  $\{P^{A_O}, P^{A_S}\}$  and  $\{P^{A_{O'}}, P^{A'_S}\}$ , but without common partition  $\{P^{A_O}, P^{A_{O'}}, P^{A_S}\}$ .

Suppose that the accessible region  $A$  of the joint system is discrete and consists of only three points:

$$\begin{aligned} a &= (x_1, y_1, z_1) \\ b &= (x_1, y_2, z_2) \\ c &= (x_2, y_2, z_3) \end{aligned}$$

where the coordinates  $x$ ,  $y$  and  $z$  refer to the microstates in  $X_O$ ,  $X_{O'}$  and  $X_S$ , respectively. Then the separate partitions are the following:

$$\begin{aligned} P^{A_O} &= \{\{x_1\}, \{x_2\}\} & P^{A_S} &= \{\{z_1, z_2\}, \{z_3\}\} \\ P^{A_{O'}} &= \{\{y_1\}, \{y_2\}\} & P'^{A_S} &= \{\{z_1\}, \{z_2, z_3\}\} \end{aligned}$$

but it is easy to see that there is no (non-trivial) common partition.

**Example 3.** A joint system  $O + O' + S$  with a common partition  $\{P^{A_O}, P^{A_{O'}}, P^{A_S}\}$ . Modify the above example as follows. Let the three microstate be the following:

$$\begin{aligned} a &= (x_1, y_1, z_1) \\ b &= (x_2, y_2, z_1) \\ c &= (x_3, y_3, z_2) \end{aligned}$$

Then the common partition is the following:

$$\begin{aligned} P^{A_O} &= \{\{x_1, x_2\}, \{x_3\}\} & P^{A_S} &= \{\{z_1\}, \{z_2\}\} \\ P^{A_{O'}} &= \{\{y_1, y_2\}, \{y_3\}\} & P'^{A_S} &= \{\{z_1\}, \{z_2\}\} \end{aligned}$$

To characterize the different possibilities we analyse Definitions 4 and 5 in terms of projective connectivity.

Consider the projection  $A_{OS}$  of the accessible region  $A$  of the joint system  $O + O' + S$  onto  $X_{OS}$ . Define the equivalence relation  $\sim$  on  $A_{OS}$  in the spirit of Definition 2. Let  $P^{A_{OS}\sim}$  denote a partition of  $A_{OS}$  if its elements are mutually projectively unconnectible in the sense defined in Section 2: for any  $x \in B$  and  $x' \in B'$  where  $B$  and  $B'$  are different elements of  $P^{A_{OS}\sim}$ ,  $x \not\sim x'$ . Denote the set of such partitions of  $A_{OS}$  by  $\mathbf{P}^{A_{OS}\sim}$ . Finally, denote by  $P_O^{A_{OS}\sim}$  and  $P_S^{A_{OS}\sim}$  the projections of  $P^{A_{OS}\sim}$  onto  $A_O$  and  $A_S$ , respectively. Let  $P^{A_{O'S}\sim}$ ,  $P_{O'}^{A_{O'S}\sim}$  and  $P_S^{A_{O'S}\sim}$  be defined similarly.

This leads to the following two definitions of macrostates which, however, are equivalent to the previous ones.

**Definition 6.** Let  $P^{A_{OS}\sim}$  and  $P^{A_{O'S}\sim}$  be partitions of  $A_{OS}$  and  $A_{O'S}$ , respectively. Then we call an element of  $P_S^{A_{OS}\sim}$  and  $P_S^{A_{O'S}\sim}$  an *unconnectibility macrostate* of the system  $S$  relative to  $O$  and  $O'$ , respectively, within the joint system  $O + O' + S$  with accessible region  $A$ .

**Definition 7.** Let  $P^{A_{OS}\sim}$  and  $P^{A_{O'S}\sim}$  be partitions of  $A_{OS}$  and  $A_{O'S}$ , respectively, such that  $P_S^{A_{OS}\sim} = P_S^{A_{O'S}\sim}$ . Then we call an element of  $P_S^{A_{OS}\sim}$  an *unconnectibility common macrostate* of the system  $S$  relative to both  $O$  and  $O'$  within the joint system  $O + O' + S$  with accessible region  $A$ .

**Proposition 2.** Definition 4 of macrostates and Definition 6 of unconnectibility macrostates are equivalent. Definition 5 of common macrostates and Definition 7 of unconnectibility common macrostates are equivalent.

The proof is a straightforward consequence of the proof of Proposition 1, therefore we omit it. We simply illustrate it on Example 2 and 3 above.

In Example 2:

$$\begin{aligned} P^{A_{OS}} &= \{ \{(x_1, z_1), (x_1, z_2)\}, \{(x_2, z_3)\} \} \\ P^{A_{O'S}} &= \{ \{(y_1, z_1)\}, \{(y_2, z_2), (y_2, z_3)\} \} \end{aligned}$$

In Example 3:

$$\begin{aligned} P^{A_{OS}} &= \{ \{(x_1, z_1), (x_2, z_1)\}, \{(x_3, z_2)\} \} \\ P^{A_{O'S}} &= \{ \{(y_1, z_1), (y_2, z_1)\}, \{(y_3, z_2)\} \} \end{aligned}$$

How do two observers take notice that their joint measurement generates a common macrostate on the system  $S$ ? Since in case of a common partition  $\mathcal{A}^{os}[P^{Ao}] = P^{As} = \mathcal{A}^{o's}[P^{A'o}]$ , therefore whenever the microstate of  $O$  is in  $B_O \in P^{Ao}$ , the microstate of  $O'$  will be in  $(\mathcal{A}^{so'} \circ \mathcal{A}^{os})(B_O) \in P^{A'o}$ . That is elements of  $P^{Ao}$  and  $P^{A'o}$  will be *perfectly correlated*. Thus, a system of common macrostates establishes a correlation between the measurement outcomes of the different observers.

Can the situation be reversed? Suppose the measurement outcomes of two observers  $O$  and  $O'$  of a system  $S$  are perfectly correlated in the sense that whenever the microstate of  $O$  is in  $B_O \in P^{Ao}$ , the microstate of  $O'$  will be in a particular  $B_{O'} \in P^{A'o}$ . Does this perfect correlation guarantee that there is a common macrostate of the system  $S$  relative to  $O$  and  $O'$ , respectively? The next example shows that even less is false.

**Example 4.** Perfect correlation between the observers does not even guarantee that there are separate macrostates of the joint system.

Suppose again that the accessible region  $A$  of the joint system is discrete and consists of the following four points:

$$\begin{aligned} a &= (x_1, y_1, z_1) \\ b &= (x_1, y_1, z_2) \\ c &= (x_2, y_2, z_1) \\ d &= (x_2, y_2, z_2) \end{aligned}$$

where again the coordinates  $x, y$  and  $z$  refer to the microstates of  $O, O'$  and  $S$ , respectively. Here the partitions

$$\begin{aligned} P^{Ao} &= \{ \{x_1\}, \{x_2\} \} \\ P^{A'o'} &= \{ \{y_1\}, \{y_2\} \} \end{aligned}$$

will perfectly correlate, but there is neither a  $\{P^{Ao}, P^{As}\}$  nor a  $\{P^{Ao'}, P^{As}\}$  (non-trivial) partition satisfying (9)-(10). Thus, perfect correlation between the measurements of the two observers does not guarantee a perfect correlation between any of the observers and the system.

Due to the absence of any correlation between the observers and the system in the above example one is not even justified to call  $O$  and  $O'$  observers of the system. It is straightforward to see however that if  $O$  is indeed an observer, that is  $O$  is perfectly correlated with  $S$  and also with  $O'$ , then  $O'$  will also do so with  $S$ : if  $\{P^{Ao}, P^{As}\}$  is a pair of partitions satisfying (9) and  $P^{Ao}$  and  $P^{Ao'}$  are perfectly correlated, then  $\{P^{Ao'}, P^{As}\}$  is a pair of partitions satisfying (10). In short,  $\{P^{Ao}, P^{Ao'}, P^{As}\}$  is a common partition of the joint system.

## 5 Distinguishability, regularity and measurement

Consider again the example of the temperature measurement. Suppose you measure the temperature of a container of gas with a thermometer. Denote the thermometer by  $O$  and the gas by  $S$ . By inserting the thermometer into the gas the experimenter introduces an interaction between the measuring device and the system. This interaction fixes which microstates are possible for the joint system  $O + S$ , that is it fixes the accessible region  $A$ . The crucial point in the coming about of macrostates is the “shape” of the accessible region. We have seen that  $A$  can theoretically be of such shape that for any microstate of  $O$  there is only one possible microstate of  $S$ . Typically, however, the accessible region is such that only *sets* of microstates of the phase spaces  $X_O$  and  $X_S$  transform mutually into one another. That is certain sets of microstates of the thermometer will correlate with certain sets of microstates of the gas in a one-to-one manner.

But how does this abstract schema relate to the real-world temperature measurement? Measuring the temperature colloquially means the reading off of the mercury level of the thermometer inserted into the gas. The different mercury levels will again generate a partition of  $X_O$  since each microstate of the thermometer uniquely determines a mercury level. But how is this partition of  $X_O$  generated by the mercury levels is related to the above partition of  $X_O$  standing in a one-to-one correlation with a partition of  $X_S$ ? This question leads us to the notion of distinguishability.

So far we have been somewhat vague about what we take to be an observer. By the same symbol  $O$  we referred to a human observer as well as to a measuring device. The reason for this ambiguity is that a direct observation of the thermometer by the sense-organs of a human observer and the measurement of the gas with a thermometer can be described in similar terms; namely, in terms of a correlation between an “observer” and a system. And this is the point where distinguishability appears. Distinguishability is nothing but an observer-system correlation such that the observer is a human observer.

Suppose an experimenter is reading off the temperature from a thermometer. The mercury levels in the glass tube are distinguishable to the naked eye while the positions of

mercury molecules are not. What does it mean that the experimenter is able to distinguish between the different mercury levels of the thermometer? Physically speaking looking at the thermometer is an interaction between the thermometer and the observer's sense-organs and brain. Although the details of this interaction are highly complex, one can still think of this interaction as a constraint on the "phase space" of the joint system composed of the observer's sense-organs and brain and the thermometer. One can then speculate that the clustering of the microstates of the thermometer into distinguishable macrostates is nothing but a one-to-one correlation between certain sets of microstates of the observer's sense-organs and brain and certain other sets of microstates of the thermometer. Being able to distinguish then physically means that certain clusterings of the state space of a physical object strongly correlate with certain clusterings of our sense organ and brain states. If this speculation is correct, then it may well be the case that the accessible region of the joint system is responsible for certain stable clusterings of our brain states and for the bringing about of the associated mental state. The short cognitive story would then be this: mental states arise where the brain can make distinctions; and making distinctions is a one-to-one correlation between coarse-grained brain states and coarse-grained states of the outer world.

According to the above description both the temperature of a gas and the lengths of the mercury column, as macrostates of the corresponding systems, are observer-relative notions. The former is defined in terms of the interaction with the thermometer, the latter is defined in terms of the interaction with the human observer. Notice that since these two kinds of macrostates refer to different interactions with different "observers", there is no a priori guarantee that the two partitions of the thermometer, namely the one generated by the distinguishable mercury levels and the one standing in a one-to-one correlation with the temperature partition of the gas, coincide. It is a contingent fact of the world that the "mercury level" reflects the temperature of the gas and at the same is time accessible for human observation.

Despite the essential observer-relativity of macrostates, there is a sense in which they are inherent to the target system in question. This sense is provided by the fact that macrostates, as subsets and partitions of the system's phase space, satisfy regularities. Consider the example of the ideal gas law. The temperature  $T$ , the pressure  $P$  and the volume  $V$  are functions on phase space. Their inverse images carve out three distinct partitions of the system's phase space corresponding to the level surfaces of these phase functions. The elements of these partitions are macrostates since they correlate with elements of the phase space partition of the appropriate measuring devices via the measurement interaction. Nevertheless, there is something inherent in the partitions generated by phase function  $T$ ,  $P$  and  $V$ : they stand in a well-defined functional relationship to one another *independently of whether they are measured or not*. Namely, if the ideal gas is in a microstate in which the temperature is  $T$  and the pressure is  $P$  then its volume  $V$  is proportional to  $1/PT$ . Again, these kinds of regularities are present regardless of being observed or not. They can be discovered, however, by an observer if she has empirical access to these partitions by various measurements. Empirical access means that

the sets of partitions featuring in thermodynamic regularities and the sets of partitions brought about by the measurements coincide or are close enough to one another. But the coincidence of the two partitions, the ones brought about by measurement and the ones featuring in thermodynamic regularities, is a contingent fact of nature, as Hemmo and Shenker rightly stress. It is contingent in the sense that the two types of partitions are determined by different physical conditions, the characteristics of the interaction between the measurement device and the system on the one hand and the characteristics of the interactions within the system on the other, and there is no a priori connection between these two types of conditions. This kind of contingency remains even if one *defines* “measurement” as a physical process that reveals inherent properties of the system satisfying certain regular connections within the system. For if one is to regard temperature measurement as a process the result of which is correlated with the value of phase function  $T$  on the system’s space – this latter now being defined merely in terms of the system’s properties, for example as the average kinetic energy of the molecules in the system –, then it becomes a contingent fact what sort of physical processes realize temperature measurement, and whether such processes exist in nature at all. One can imagine a possible world where the functional relationship of phase functions  $T$ ,  $P$  and  $V$ , as inherent properties of a system, are the same as in the actual world, but where those physical principles on which experiencing and measuring temperature are based – heat transfer, equalization of temperature, blackbody radiation, temperature dependence of electrical resistance, etc. – are different or do not exist at all. If these did not even exist in such a possible world then, despite the existence of a “thermodynamic regularity” between  $T$  and other variables on phase space, temperature wouldn’t feature as a macrostate in the sense discussed in our paper.

It is worth noting that despite their conceptual difference the emergence of partitions brought about by interaction with an observer and partitions brought about by regularities internal to the system can be described in the same mathematical framework. The essential idea common to both concepts can be captured as the correlation, regularity or functional relationship between the values of certain functions defined on a system’s phase space (where in the observer-relative case by “system” we mean the joint system of the target system and the observer). The reason is that the notion of thermodynamic regularity described as a functional relationship between phase functions (for example as the one between  $T$ ,  $P$  and  $V$ ) is essentially the same notion as the one corresponding to our weakest definition of macrostate in Section 2, Figure 5, with the only difference that the system’s phase space should be substituted with the joint phase space of the target system and the observer (and functions  $f$  and  $g$  on Figure 5 are thought to be defined on the whole joint phase space rather than on the axes). The main idea is that both notions can be cashed out in terms of the most general description of *regularity* either within the system’s or within the observer plus system’s state space.

Finally, let us make a general observation on the observer-relativeness of macrostates investigated formally in the previous section. Suppose that we measure the temperature of a container of gas with a mercury thermometer  $O$  and by an alcohol thermometer  $O'$ .

Measuring the temperature with a mercury thermometer generates a partition of  $X_S$  and thus brings about macrostates relative to  $O$ , as explicated earlier. Measuring the temperature with an alcohol thermometer brings about macrostates relative to  $O'$ . There is no a priori guarantee that these two partitions have anything in common. However, in practice we observe a one-to-one correlation between the readings of the mercury thermometer and the alcohol thermometer. As we saw in the end of the previous section, whenever  $O$  and  $O'$  are perfectly correlated, and  $O$  and  $O'$  are observers of a system  $S$  in the sense of there being correlation between  $O$  and  $S$  and between  $O'$  and  $S$ , then there always exists a system of common macrostates relative to  $O$  and  $O'$ . This common macrostate system can be regarded as a common causal explanation of the perfect correlation of measurements  $O$  and  $O'$ . It is the temperature of the gas, an inherent property of the system pertaining to a partition of its phase space, that explains the correlation of the two kinds of “temperature” measurement.

## 6 Conclusions

In this paper we defined macrostates, following Hemmo and Shenker, as elements of certain partitions of the system’s phase space generated by the interaction between the system and an observer. We investigated several formal and conceptual features of this notion and proved the equivalence of two different characterizations of macrostates. Finally, we intended to accommodate the other two features of macrostates, namely thermodynamic regularity and distinguishability within this framework.

## Appendix

### Proof of Proposition 1.

We prove Proposition 1 via proving four lemmas.

**Lemma 1.**  $P_O^{A\sim}$  and  $P_S^{A\sim}$  are partitions of  $A_O$  and  $A_S$ , respectively, that is  $P_O^{A\sim} \in \mathbf{P}^{A_O}$  and  $P_S^{A\sim} \in \mathbf{P}^{A_S}$ .

*Proof.* We show that the sets of  $P_O^{A\sim}$  are disjoint and add up to  $A_O$ . (For  $P_S^{A\sim}$  the proof is similar.)

Suppose that sets of  $P_O^{A\sim}$  are *not* disjoint, that is there exists a  $b \in (B_O \cap B'_O)$  with  $B$  and  $B'$  being different elements in  $P_O^{A\sim}$ . Then there exist an  $x \in B$  and an  $x' \in B'$  such that  $x_O = x'_O = b$ . But then  $x \sim x'$ . Contradiction.

Suppose that sets of  $P_O^{A\sim}$  are *not* adding up to  $A_O$ , that is there exists a  $b \in A_O$  such that  $b \notin B_O$  for any  $B_O \in P_O^{A\sim}$ . Then  $\mathcal{A}^O(b) \cap B = \emptyset$  for any  $B \in P_O^{A\sim}$ , that is  $P_O^{A\sim}$  is not a partition of  $A$ . Contradiction. ■

**Lemma 2.** The partition  $P^{A\sim}$  can be “reconstructed” from its projections in the sense that  $\mathcal{A}^O[P_O^{A\sim}] = \mathcal{A}^S[P_S^{A\sim}] = P^{A\sim}$ .



*Proof.* We show that  $\mathcal{A}^o[P_O^{A\sim}] = P^{A\sim}$ . (For  $\mathcal{A}^s[P_S^{A\sim}] = P^{A\sim}$  the proof is similar.) Suppose to the contrary that  $\mathcal{A}^o[P_O^{A\sim}] \neq P^{A\sim}$ . This means that there exist an  $x \in A$  and a  $B \in P^{A\sim}$  such that either (i)  $x \in \mathcal{A}^o[B_O]$  and  $x \notin B$ , or (ii)  $x \notin \mathcal{A}^o[B_O]$  and  $x \in B$ .

As for case (i), since  $x \in \mathcal{A}^o[B_O] \setminus B$  and therefore  $x_O \in B_O$ , there exists an  $x' \in B$  such that  $x'_O = x_O$ . But then  $x \sim x'$  and hence  $x \in B$ . Contradiction.

Case (ii) can be excluded since for any region  $B$  in  $A$ ,  $B \subseteq \mathcal{A}^o[B_O]$ . ■

The following Lemma is a straightforward corollary of Lemma 2.

**Lemma 3.**  $\mathcal{A}^{os}[P_O^{A\sim}] = P_S^{A\sim}$  and  $\mathcal{A}^{so}[P_S^{A\sim}] = P_O^{A\sim}$ .

*Proof.*  $\mathcal{A}^{os}[P_O^{A\sim}] = (\mathcal{A}^o[P_O^{A\sim}])_S = (P^{A\sim})_S = P_S^{A\sim}$  and similarly for  $\mathcal{A}^{so}[P_S^{A\sim}]$ . ■

Obviously, Lemma 3 shows that macrostates according to Definition 3 are macrostates also according to Definition 1 (and hence also according to Definitions ??-??). Our next Lemma demonstrates that the converse is also true.

**Lemma 4.** Suppose that  $\{P^{A_O}, P^{A_S}\}$  are partitions of  $A_O$  and  $A_S$ , respectively, such that  $\mathcal{A}^{os}[P^{A_O}] = P^{A_S}$  and  $\mathcal{A}^{so}[P^{A_S}] = P^{A_O}$ . Then there exists a  $P^{A\sim} \in \mathbf{P}^{A\sim}$  such that  $P^{A_O} = P_O^{A\sim}$  and  $P^{A_S} = P_S^{A\sim}$ .

*Proof.* (i) First we show that  $\mathcal{A}^o[P^{A_O}]$  is a partition of  $A$ , that is  $\mathcal{A}^o[P^{A_O}] \in \mathbf{P}^A$ . Suppose that sets of  $\mathcal{A}^o[P^{A_O}]$  are *not* disjoint, that is for a given  $B_O, B'_O \in P^{A_O}$  there exists an  $x$  such that  $x \in (\mathcal{A}^o[B_O] \cap \mathcal{A}^o[B'_O])$ . But then  $x_O \in (B_O \cap B'_O)$ , that is  $P^{A_O}$  is not a partition of  $A_O$ . Contradiction.

Suppose that sets of  $\mathcal{A}^o[P^{A_O}]$  are *not* adding up to  $A$ , that is there exists an  $x$  in  $A$  such that  $x \notin B$  for any  $B \in \mathcal{A}^o[P^{A_O}]$ . Then  $x_O \notin B_O$  for any  $B_O \in P^{A_O}$ , that is  $P^{A_O}$  is again not a partition of  $A_O$ . Contradiction.

Hence,  $\mathcal{A}^o[P^{A_O}]$  is a partition of  $A$ . Similarly,  $\mathcal{A}^s[P^{A_S}]$  is a partition of  $A$ .

(ii) Next we show that  $\mathcal{A}^o[P^{A_O}] = \mathcal{A}^s[P^{A_S}]$ .

Let  $B_O \in P^{A_O}$  and let  $B_S$  be its corresponding element  $B_S = \mathcal{A}^{os}[B_O]$ . Suppose that  $\exists x \in \mathcal{A}^s[B_S]$  such that  $x \notin \mathcal{A}^o[B_O]$ . Then  $x_O \notin B_O$  and hence  $x_O \in B'_O$  for some  $B'_O \neq B_O$ ,  $B'_O \in P^{A_O}$ . But since  $\mathcal{A}^{os}[B'_O] = B'_S$  for some  $B'_S \neq B_S$ ,  $B'_S \in P^{A_S}$ , we would have  $x_S \in B'_S$  and thus  $x \in \mathcal{A}^s[B'_S]$  where  $\mathcal{A}^s[B'_S] \neq \mathcal{A}^s[B_S]$ , a contradiction. The argument is similar for the case when  $\exists x \in \mathcal{A}^o[B_O]$  such that  $x \notin \mathcal{A}^s[B_S]$ . Hence for all  $B_O \in P^{A_O}$  and for their corresponding  $B_S \in \mathcal{A}^{os}[B_O]$  we have  $\mathcal{A}^o[B_O] = \mathcal{A}^s[B_S]$ , and thus  $\mathcal{A}^o[P^{A_O}] = \mathcal{A}^s[P^{A_S}]$ .

(iii) Finally, we show that  $\mathcal{A}^o[P^{A_O}] = \mathcal{A}^s[P^{A_S}] \in \mathbf{P}^{A\sim}$ .

Suppose that  $\mathcal{A}^o[P^{A_O}] \notin \mathbf{P}^{A\sim}$ , that is there exists an  $x \in B$  and an  $x' \in B'$  such that  $x \sim x'$  and  $B$  and  $B'$  are distinct elements of  $\mathcal{A}^o[P^{A_O}]$ . Let  $\{x_n\}_{n=1}^N$  be the sequence connecting  $x$  and  $x'$  and suppose that  $(x_2)_S = x_S$  (the argument is similar when  $(x_2)_O = x_O$ ). Now,  $x_2$  cannot be in a  $B_2 \in \mathcal{A}^o[P^{A_O}]$  such that  $B_2 \neq B$ , otherwise  $(x_2)_S$  were in  $B_S \cap (B_2)_S$  and consequently  $B_S$  and  $(B_2)_S$  were not disjoint. So  $x_2 \in B$ . By induction, we obtain that  $x_3, \dots, x_N = x'$  are all in  $B$ . Hence  $B$  and  $B'$  are not distinct. Contradiction. ■

By this we also complete the proof of Proposition 1. ■

**Acknowledgements.** This work has been supported by the Hungarian Scientific Research Fund, OTKA K-115593 and by the Bilateral Mobility Grant of the Hungarian and Polish Academies of Sciences, NM-104/2014.

## References

Hemmo, M. and O. Shenker. *The Road to Maxwell's Demon, Conceptual Foundations of Statistical Mechanics*. Cambridge: Cambridge University Press, 2012.