AXIOMATIC TESTS FOR THE BOLTZMANN DISTRIBUTION

SIMONE CERREIA-VIOGLIO, FABIO MACCHERONI, MASSIMO MARINACCI,
AND ALDO RUSTICHINI

Abstract. The Boltzmann distribution family describes a single parameter (temperature) class of probability distributions over a state space; at any given temperature, the ratio of probabilities of two states depends exponentially on their difference in energy. Beyond physics, this distribution family is very popular in many important disciplines, under different names with different interpretations.

Such widespread use in diverse fields suggests a common conceptual structure. We identify such a structure on the basis of few natural axioms that can be statistically tested. Our axiomatic characterization thus provides alternative empirical tests of the Boltzmannian modeling theories.

1. Introduction

According to the classic *Boltzmann distribution* of statistical mechanics, when the energy associated with some state a of a system is $E(a)$, the frequency with which that state occurs in equilibrium is proportional to

$$e^{-\frac{E(a)}{kt}}$$

where t is the system absolute temperature and k is the Boltzmann constant.

Under different interpretations and names (e.g., *softmax* or *Multinomial Logit*), the Boltzmann distribution is widely used in many fields of science, from physics to computer science, from economics to psychology. For example, in economics the Multinomial Logit distribution is the workhorse of discrete choice analysis. It gives the probability that an agent with a utility function $V = -E$ selects an alternative a when trying to maximize V but, say because of lack of information, makes mistakes in evaluating the various alternatives. In this case, the standard deviation of mistakes is proportional to t, that is, temperature is replaced by noise.[1] More recently, in econophysics the Boltzmann distribution has been used to describe market imperfections (with E representing the bid-ask spreads of quotations) and income distributions (with E representing the amounts of money corresponding to wealth levels).[2]

In this paper, we provide an axiomatic characterization of the Boltzmann distribution based on observables. Specifically, we show that a family $p = \{p_t\}$ of conditional distributions

---

We wish to thank Carlo Baldassi, Daniele Durante, Claudio Tebaldi, Piero Veronese, Riccardo Zecchina, the editor and an anonymous referee for very helpful comments, as well as Renato Berlinghieri, Marco Pirazzini, and Giulio Principi for brilliant research assistance. Part of the material of this paper was first circulated in the IGIER working paper 593 of 2016.

[1] See Train [16] for a textbook presentation. Later we will discuss another recent use of the Multinomial Logit distribution in economics (Section 7.1).

[2] See the letter of Kanazawa et al. [8], and the colloquium of Yakovenko and Rosser [18].

satisfies certain properties if and only if there exists an energy function E such that

$$p_t(a \mid A) = \frac{e^{-\frac{E(a)}{kt}}}{\sum_{b \in A} e^{-\frac{E(b)}{kt}}}$$

for all temperatures t and all states a in a collection A of accessible states. The function E is unique up to an additive constant and can be retrieved from data. Besides a common conceptual structure for this ubiquitous distribution, our axiomatic analysis thus also provides new empirical tests for it. Differently from the existing characterizations of the Boltzmann distribution, in ours the energy function E is derived rather than assumed.[3]

2. Notation

We denote by $\mathcal{A}$ the collection of all finite subsets A of a universal system of states X, with $|X| \geq 2$, and by p a *random state function*

$$\begin{array}{rccc} p: & (0,\infty) \times X \times \mathcal{A} & \to & \mathbb{R}_+ \\ & (t, a, A) & \mapsto & p_t(a \mid A) \end{array}$$

that associates to a triplet (t, a, A) the frequency $p_t(a \mid A)$ of state $a \in X$, at temperature t when A is the subsystem of accessible states.

Clearly,

$$p_t(B \mid A) = \sum_{b \in B} p_t(b \mid A)$$

is the conditional frequency of some state in $B \subseteq A$. For a binary subsystem, we just write $p_t(a, b) = p_t(a \mid \{a, b\})$ for the frequency of a state a, with its odds denoted by

$$r_t(a, b) = \frac{p_t(a, b)}{p_t(b, a)}$$

Finally, δ_a is the point mass at $a \in X$, i.e., $\delta_a(A) = 1$ if $a \in A$ and $\delta_a(A) = 0$ otherwise.

3. Axioms and results

We consider a few axioms on a given random state function $p : (0, \infty) \times X \times \mathcal{A} \to \mathbb{R}_+$ that describes the statistical behavior of the system. We begin with positivity and conditioning axioms that require each section p_t of p to be a conditional probability system (see Renyi [13] and Luce [11]).

Axiom A. 1 (Positivity). *Given any* $(t, A) \in (0, \infty) \times \mathcal{A}$,

$$\sum_{a \in X} p_t(a \mid A) = 1$$

with $p_t(a \mid A) > 0$ *if and only if* $a \in A$.

Axiom A. 2 (Conditioning). *Given any* $(t, A) \in (0, \infty) \times \mathcal{A}$,

$$p_t(b \mid A) = p_t(b \mid B)\, p_t(B \mid A)$$

for all $B \subseteq A$ *and all* $b \in B$.

[3] See Section 5 for details and for a comparison with other axiomatic approaches in physics to the Boltzmann distribution.

The next axiom requires the conditional probability systems p_t to vary continuously with temperature.

Axiom A. 3 (Continuity)**.** *Given any* $(a, A) \in X \times \mathcal{A}$*,*

$$\lim_{t \to s} p_t (a \mid A)$$

exists for all $s \geq 0$ *and coincides with* $p_s (a \mid A)$ *when* $s > 0$*.*

Continuity guarantees, *inter alia*, that, as t goes to 0, a limit probability $p_0 (a \mid A)$ is defined for all $(a, A) \in X \times \mathcal{A}$. The following axiom requires the consistency of freezing and positive temperature probabilities.

Axiom A. 4 (Consistency)**.** *Given any* $a, b \in X$*,*

$$p_t (a, b) > p_t (b, a) \implies p_0 (a, b) > p_0 (b, a)$$

for all $t > 0$*.*

Next we postulate that, if at a zero temperature a binary subsystem is not deterministically in either state, then both states are equally likely.

Axiom A. 5 (Zero Uniformity)**.** *Given any* $a, b \in X$*,*

$$p_0 (a, b) \neq 0, 1 \implies p_0 (a, b) = 1/2$$

Our final axiom ties together the conditional distributions at different temperatures. It requires the dependence of odds from inverse temperature not to be infinitely far from exponential. It is just a "grain of exponentiality" in the evolution of the system that, as our next theorem shows, develops into precisely an exponential dependence of odds on inverse temperatures.

Axiom A. 6 (Boundedness)**.** *Given any* $a, b \in X$*,*

$$\sup_{t,s \in (0,\infty)} \left| r_{\frac{1}{t+s}} (a, b) - r_{\frac{1}{t}} (a, b) \, r_{\frac{1}{s}} (a, b) \right| < \infty$$

We can now state our first result, in which we characterize the Boltzmann distribution.

Theorem 1. *A random state function* $p : (0, \infty) \times X \times \mathcal{A} \to \mathbb{R}_+$ *satisfies A.1–A.6 if and only if there exists a function* $E : X \to \mathbb{R}$ *such that*

$$p_t (a \mid A) = \frac{e^{-\frac{E(a)}{kt}}}{\sum_{b \in A} e^{-\frac{E(b)}{kt}}} \delta_a (A) \tag{1}$$

for all $(t, a, A) \in (0, \infty) \times X \times \mathcal{A}$*.*

In this case, the function E *is unique up to an additive constant.*

In view of this result, it is natural to say that a random state function p is *Boltzmannian* if it satisfies A.1–A.6. A natural question is whether one can replace the thermal energy kt with a more general noise term $\kappa (t)$. To address this question, we introduce a generic binary operation, concatenation, written $\oplus$, that has the usual sum $+$ as a special case.

Definition 1. *A* concatenation *is a binary operation* $\oplus$ *on* $\mathbb{R}_+$ *which is associative, commutative, with identity element* 0*, and such that*

$$t > s \implies t \oplus v > s \oplus v \qquad \forall v \in (0, \infty)$$

Besides the sum, other simple examples of concatenation are $t \oplus s = t + s + \eta ts$ and $t \oplus s = \sqrt[\eta]{t^\eta + s^\eta}$ for some $\eta \in (0, \infty)$.

The next axiom is the obvious extension of A.6 to a generic concatenation. It continues to have a "grain of exponentiality" nature.

Axiom A. 7 (Weak Boundedness)**.** *Given any* $a, b \in X$*,*

$$\sup_{t,s \in (0,\infty)} \left| r_{\frac{1}{t \oplus s}}(a, b) - r_{\frac{1}{t}}(a, b)\, r_{\frac{1}{s}}(a, b) \right| < \infty$$

for a continuous concatenation $\oplus$*.*

We can now generalize the earlier Boltzmannian result, which is the special case of the theorem below when the concatenation $\oplus$ is the usual sum $+$. A final notion is needed: p is *uniform* when $p_t(a \mid A) = \delta_a(A) / |A|$ for all $(t, a, A) \in (0, \infty) \times X \times \mathcal{A}$.

Theorem 2. *A random state function* $p : (0, \infty) \times X \times \mathcal{A} \to \mathbb{R}_+$ *satisfies A.1–A.5 and A.7 if and only if there exist a function* $E : X \to \mathbb{R}$ *and an increasing bijection* $\kappa : (0, \infty) \to (0, \infty)$ *such that*

$$p_t(a \mid A) = \begin{cases} \dfrac{e^{-\frac{E(a)}{\kappa(t)}}}{\sum_{b \in A} e^{-\frac{E(b)}{\kappa(t)}}} & a \in A \\ 0 & a \notin A \end{cases} \tag{2}$$

for all $(t, a, A) \in (0, \infty) \times X \times \mathcal{A}$*.*

In this case, p *is uniform if and only if* E *is a constant function. When* p *is not uniform:*

(i) *functions* $\tilde{E}$ *and* $\tilde{\kappa}$ *also represent* p *as in (2) if and only if there exist* $m > 0$ *and* $q \in \mathbb{R}$ *such that* $\tilde{E} = mE + q$ *and* $\tilde{\kappa} = m\kappa$*;*

(ii) *the only concatenation* $\oplus$ *for which A.7 holds is*

$$t \oplus s = \phi^{-1}[\phi(t) + \phi(s)] \qquad \forall t, s \in [0, \infty)$$

where $\phi : [0, \infty) \to [0, \infty)$ *is given by* $\phi(v) = 1/\kappa(1/v)$ *for all* $v > 0$ *and* $\phi(0) = 0$*.*

4. Convex energy

The physical question that the Boltzmann distribution addressed was: What is the distribution of velocities in a gas at a certain temperature? The space of states (velocities) is a convex set and energy, which is proportional to square speed, is a convex function. Analogously, in economics concave utility functions play a fundamental role.

This motivates the next result that characterizes convex energy (so, concave utility).

Proposition 3. *Let* X *be a convex set and* p *a Boltzmannian random state function with energy* E*. The following conditions are equivalent:*

(i) *the function* $E : X \to \mathbb{R}$ *is convex;*

(ii) *there exists* $t \in (0, \infty)$ *such that*

$$p_{\alpha t}(\alpha a + (1-\alpha) b, b) \geq p_t(a, b) \tag{3}$$

for all $a, b \in X$ *and all* $\alpha \in (0, 1)$*;*

(iii) *given any* $s \in (0, \infty)$*,*

$$p_s\left(b \;\middle|\; \frac{1}{\eta}A + \left(1 - \frac{1}{\eta}\right) b\right) \leq p_{\eta s}(b \mid A) \tag{4}$$

for all $A \in \mathcal{A}$*, all* $b \in A$*, and all* $\eta > 1$*;*

(iv) *given any* $s \in (0, \infty)$*,*

$$p_s\left(b \;\middle|\; \frac{1}{\eta}A + \left(1 - \frac{1}{\eta}\right) b\right) \leq p_{\eta s}(b \mid A) \tag{5}$$

for all $A \in \mathcal{A}$*, all* $b \in \arg\min_{a \in A} p_{\eta s}(a \mid A)$*, and all* $\eta > 1$.

The stochastic choice interpretation of this result is based on the trade-off between noise (temperature) and states' distinguishability. By mixing states we make them closer, so less distinguishable, and we augment the probability of making a mistake. But the frequency of mistakes also augments as noise (temperature) increases. Proposition 3 says that utility is concave (energy is convex) if and only if decreasing distinguishability by a factor $\eta > 1$ is less detrimental than increasing noise by the same factor.

5. Axiomatizations in Physics

Boltzmann himself derived, in 1877, the eponymous distribution via an axiomatic strategy that consisted in finding the *most likely macrostate for given total energy under the assumption of uniform probability of accessible microstates*. This is clear in the translation of the original paper, with a scientific commentary, by Sharp and Matschinsky [15] and in the extensive treatment of Bowler [4].

In another seminal axiomatization, Jaynes [7] showed that the Boltzmann distribution is *the maximizer of entropy subject to the constraint given by average energy*. In this way, the Boltzmann distribution appears as the least biased distribution compatible with a given average energy value. This approach has inspired some important contributions in theoretical economics, for example the costly information acquisition analysis of choice under uncertainty of Matejka and McKay [12] (see below).

Further axiomatizations build on *independence of configurations for conserved sum of energies*. These approaches lead to an exponential functional equation in the energy variable. See, e.g., Landau and Lifshitz [9], Eisberg and Resnick [6], and again Bowler [4].

In all these three (classes of) axiomatizations the total amount of *energy is given* and so is the energy function. Moreover, in the last approach the dependence of the probability of a given state on its energy level is explicitly postulated. In contrast, in our approach the *energy function is derived*. Our axioms provide a test for the existence of an energy function $E : X \to \mathbb{R}$ such that, for every value t of temperature, the probabilities of states depend on their energy via equation (1). In other words, our axioms allow to verify Boltzmannian modeling theories without prior knowledge of the energy function. On the technical side, this is achieved by imposing a mild requirement of exponentiality relative to inverse temperatures (Axiom A.6) rather than relative to energy levels —like in the third approach described above.

6. Testing Boltzmann Theories

In this section we examine how our axiomatic analysis can be used to test whether a random state function is Boltzmannian and, if this is the case, to identify the energy function.

6.1. **Alternative axioms and identification of the parameters.**

We start with convenient reformulations of Axioms A.4 and A.7.

Axiom A. 8 (Monotonicity)**.** *Given any* $a, b \in X$, $\lim_{v \to \infty} r_v(a, b) = 1$*; moreover,*

$$r_t(a, b) > 1 \iff r_s(a, b) > r_t(a, b)$$

for all $s < t$ *in* $(0, \infty)$.

Axiom A. 9 (Concatenation)**.** *Given any* $a, b, c, d \in X$,

$$r_v(a, b) > r_t(a, b) r_s(a, b) \implies r_v(c, d) > r_t(c, d) r_s(c, d)$$

for all $s, t, v \in (0, \infty)$ *such that* $r_v(a, b) > 1$ *and* $r_v(c, d) > 1$.

These versions permit a reformulation of Theorem 2 that shows how the energy and noise functions can be directly derived from the random state function.

Theorem 4. *A random state function* $p : (0, \infty) \times X \times \mathcal{A} \to \mathbb{R}_+$ *satisfies A.1–A.3, A.5, A.8, and A.9 if and only if there exist a function* $E : X \to \mathbb{R}$ *and an increasing bijection* $\kappa : (0, \infty) \to (0, \infty)$ *such that*

$$p_t(a \mid A) = \begin{cases} \dfrac{e^{-\frac{E(a)}{\kappa(t)}}}{\sum_{b \in A} e^{-\frac{E(b)}{\kappa(t)}}} & a \in A \\ 0 & a \notin A \end{cases} \tag{6}$$

for all $(t, a, A) \in (0, \infty) \times X \times \mathcal{A}$.

In this case, p *is uniform if and only if* E *is constant. When* p *is not uniform, by arbitrarily choosing* $v \in (0, \infty)$ *and* $c, d \in X$ *such that* $p_v(c, d) > p_v(d, c)$*, then (6) holds for* $E : X \to \mathbb{R}$ *and* $\kappa : (0, \infty) \to (0, \infty)$ *defined by*

$$E(a) = \ln r_v(c, a) \qquad \text{and} \qquad \kappa(t) = \frac{\ln r_v(c, d)}{\ln r_t(c, d)} \tag{7}$$

for all $(t, a) \in (0, \infty) \times X$.

The expressions in (7) show that with large amounts of data it is possible to learn E and κ by computing the asymptotic odds $r_t(a, b)$ for all triplets (t, a, b).

The identification of $\kappa(t)$ given by (7) also implies that the Boltzmannian theories we are after, in which $\kappa(t) = kt$, correspond to a constant $t \ln r_t(c, d)$. This suggests the following powerful axiom.

Axiom A. 10 (Inversion)**.** *Given any $a, b \in X$,*

$$t \ln r_t(a,b) = s \ln r_s(a,b)$$

for all $s, t \in (0, \infty)$.

This axiom is particularly amenable to statistical testing because it replaces Axioms A.3–A.6 with a single equation.[4] Hence the relevance of the next corollary.

Corollary 5. *A random state function $p : (0,\infty) \times X \times \mathcal{A} \to \mathbb{R}_+$ satisfies A.1, A.2, and A.10 if and only if there exists a function $E : X \to \mathbb{R}$ such that*

$$p_t(a \mid A) = \frac{e^{-\frac{E(a)}{kt}}}{\sum_{b\in A} e^{-\frac{E(b)}{kt}}} \delta_a(A) \tag{8}$$

for all $(t, a, A) \in (0,\infty) \times X \times \mathcal{A}$.

In this case, arbitrarily choosing $s \in (0,\infty)$ and $c \in X$, expression (8) holds for $E : X \to \mathbb{R}$ given by

$$E(a) = ks \ln r_s(c, a) \tag{9}$$

for all $a \in X$.

6.2. **The null hypothesis.** To exemplify the statistical test of our axioms, we assume that X is finite and replace $(0,\infty)$ with a finite (ideally large) set of temperatures T.[5] With this, a positive random state function can be seen as a vector of probability measures

$$p = \left[p_t(\cdot \mid A) \in \Delta^+(A) : (t, A) \in T \times \mathcal{A}\right]$$

where $\Delta^+(A)$ denotes the relative interior of the simplex in $\mathbb{R}^A_+$. Simple combinatorics shows that the dimension of this vector is $\ell = |T| \times |X| \, 2^{|X|-1}$. Thus, setting $\Theta_{t,A} = \Delta^+(A)$ for all $(t, A) \in T \times \mathcal{A}$, it follows that the set of all random state functions is

$$\Theta = \prod_{(t,A)\in T\times\mathcal{A}} \Theta_{t,A} \subseteq \mathbb{R}^\ell$$

Each Axiom A.i (for i $= 1, 2, ..., 10$) requires the "true state function p" to belong to a subset Θ^{i} of Θ. Testing Axiom A.i then means testing the statistical hypothesis

$$H_0^{\mathrm{i}} : p \in \Theta^{\mathrm{i}} \quad \text{versus} \quad H_1^{\mathrm{i}} : p \notin \Theta^{\mathrm{i}}$$

This is a classical task in statistics that can be accomplished with both parametric approaches —e.g., the likelihood ratio or the Bayes factor test— and non parametric ones —e.g., the Kolmogorov-Smirnov or the Cramer-von Mises test.[6]

In light of Corollary 5, two sharp axioms to test are A.2 and A.10. Next we characterize the corresponding hypotheses H_0^2 and H_0^{10}.[7]

[4]The drawback of this axiom for decision theoretic applications is that its behavioral meaning is quite obscure, while the interpretation of Axioms A.3–A.6 is more natural.

[5]This reduction is not conceptually necessary, but guarantees finite dimensionality of the set of parameters. At the same time, it is not conceptually problematic, because Corollary 5 continues to hold for any set T of temperatures (see the final part of its proof).

[6]See Lehmann and Romano [10] for a textbook presentation of hypothesis testing.

[7]A similar analysis can be performed for the other axioms, we focus on A.2 and A.10 for illustrative purposes. Note that we assume positivity throughout (its corresponding Θ^1 and H_0^1 are obvious).

6.2.1. *Characterization of* Θ^2. Note that θ satisfies A.2 if and only if

$$\theta_t(b \mid A) = \theta_t(b \mid B)\,\theta_t(B \mid A)$$

for all $b \in B \subseteq A \in \mathcal{A}$ and all $t \in T$. By simple computations, this is equivalent to

$$\frac{\theta_{t,A}(a)}{\theta_{t,A}(b)} = \frac{\theta_{t,\{a,b\}}(a)}{\theta_{t,\{a,b\}}(b)}$$

for all $a, b \in A \in \mathcal{A}$ and all $t \in T$, where $\theta_{t,A}(a)$ is the a-th component of the vector $\theta_{t,A} \in \mathbb{R}_+^A$ which in turn is the (t, A)-th component of the vector θ. Thus,

$$\Theta^2 = \left\{\theta \in \Theta : \frac{\theta_{t,A}(a)}{\theta_{t,A}(b)} = \frac{\theta_{t,\{a,b\}}(a)}{\theta_{t,\{a,b\}}(b)} \quad \forall (t, a, b, A) \in T \times A^2 \times \mathcal{A}\right\}$$

Hypothesis $H_0^2 : \theta \in \Theta^2$ can be rejected when states have different degrees of similarity, this fact is well discussed in the discrete choice analysis literature (see, e.g., Chapter 3 of Train [16]), and we present a novel example in Appendix C.

6.2.2. *Characterization of* Θ^{10}. Note that θ satisfies A.10 if and only if

$$t \ln \frac{\theta_{t,\{a,b\}}(a)}{\theta_{t,\{a,b\}}(b)} = s \ln \frac{\theta_{s,\{a,b\}}(a)}{\theta_{s,\{a,b\}}(b)}$$

for all $a, b \in X$ and all $s, t \in T$. Thus,

$$\Theta^{10} = \left\{\theta \in \Theta : \left(\frac{\theta_{s,\{a,b\}}(a)}{\theta_{s,\{a,b\}}(b)}\right)^{\frac{s}{t}} = \frac{\theta_{t,\{a,b\}}(a)}{\theta_{t,\{a,b\}}(b)} \quad \forall (s, t, a, b) \in T^2 \times X^2\right\}$$

Hypothesis $H_0^{10} : \theta \in \Theta^{10}$ is typically rejected in the generalized Boltzmannian case (6) when the noise function $\kappa(t)$ is not linear.

6.3. **Testing with data.** In view of the previous discussion, a statistical test of the Boltzmann theory *that does not require an explicit expression of the probability distribution under scrutiny* is

$$H_0 : p \in \Theta_0 \qquad \text{versus} \qquad H_1 : p \notin \Theta_0 \tag{10}$$

where $\Theta_0 = \Theta^2 \cap \Theta^{10}$.

This test can be performed in many ways. Yet, the specific choice of the test and its analytical/computational implementation goes beyond the scope of this paper and inevitably depends on the data collection technique. For this reason, we just provide an elementary example as a proof of concept. A *data set* is a collection

$$D = \{(a_i, t_i, A_i) : i = 1, \ldots, I\}$$

where $(a_i, t_i, A_i) \in X \times T \times \mathcal{A}$ and $a_i \in A_i$ for each i. The probability of observing D, given $\theta \in \Theta$, is

$$L_D(\theta) = \prod_{i=1}^{I} \theta_{t_i}(a_i \mid A_i)$$

and the *likelihood ratio statistic* for (10) —that is, for testing Axioms A.2 and A.10 jointly—is

$$\lambda(D) = \frac{\sup_{\theta \in \Theta_0} L_D(\theta)}{\sup_{\theta \in \Theta} L_D(\theta)} \tag{11}$$

Once a critical threshold c is adopted, say $c = 0.99$, the null hypothesis H_0 (the Boltzmann theory) is rejected if $\lambda(D) \leq c$, otherwise it is accepted.

Summing up, thanks to their axiomatic foundation, our tests of Boltzmannian theories do not require any prior knowledge of the energy function. As a by-product, the testing procedure itself delivers a maximum likelihood estimate of the energy function. Indeed, the $\hat{\theta}$ that maximizes the numerator of (11) is the maximum likelihood Boltzmannian random state function, given data D. Our identification results (Theorem 4 and Corollary 5) then allow to derive from $\hat{\theta}$ a maximum likelihood estimate $\hat{E}$ of the energy function.

7. Additional remarks

7.1. **Optimal information acquisition.** In economics, the Multinomial Logit distribution has been used to formalize versions of the discovered preference hypothesis, where the utility function $V = -E$ is to be learned by an agent who confronts a cost t of acquiring and processing one unit of information. In particular, Matejka and McKay [12] showed that the Multinomial Logit distribution gives the optimal choice probability with which such an agent chooses an alternative a from a set A of (a priori homogeneous) available alternatives. Our axioms allow an analyst who controls t to test this theory.[8]

Remarkably, in this case axiom A.10 has a neat psychophysics interpretation (see Luce [11]). It says that, given any two alternatives a and b, the quantity

$$t \ln r_t(a, b)$$

remains constant. That is, the strength $\ln r_t(a, b)$ of the revealed preference for a over b is inversely proportional to the cost of information acquisition t. The smaller this cost is, the more information is gathered and so the better are understood the relative values of alternatives. Until, in the limit, the best alternative is chosen without error.

In this economic setting, the concavity of the utility function is based on the trade-off between information cost and alternatives' distinguishability. Under this interpretation, Proposition 3 says that utility is concave if and only if decreasing distinguishability by a factor $\eta > 1$ is less detrimental than increasing information cost by the same factor.

7.2. **Tsallis distributions.** The Boltzmann distribution with inverse temperature β, obtained by minimization of Shannon entropy, is easily seen to be the limit for $q \to 1$ of the distributions obtained by minimizing the Tsallis q-entropy. So, a natural development of this paper would consist in characterizing, for every $q \in \mathbb{R}$, the family of distributions

$$p_\beta^q(a \mid A) = \frac{(1 - \beta(q-1)E(a))^{\frac{1}{q-1}}}{\sum_{b \in A}(1 - \beta(q-1)E(b))^{\frac{1}{q-1}}}$$

[8] Different sets of axioms for more general Multinomial Logit forms appear in the subsequent papers of Saito [14] and Cerreia-Vioglio et al. [5] .

for all inverse temperatures β and all states a in a collection A of accessible states (see Tsallis [17]). This is the object of current investigation.

8. Conclusions

In this paper we provide some novel axiomatic characterizations of the Boltzmann distribution based on observables. Differently from the characterizations of this distribution existing in the physics literature, in ours the energy and noise functions are derived rather than assumed. Moreover, we characterize axiomatically the convexity of the energy function and we present the basic techniques that permit to test statistically the axioms through empirical data.

Our exercise has two objectives. First, axioms provide a principled approach that makes transparent the conceptual assumptions of a theory. Second, when regarded as "statistical hypotheses," axioms make theories testable through experiments and permit to identify their parameters. Both these objectives are here especially important because of the great influence of Boltzmaniann theories in natural and social sciences.

References

[1] J. Aczel (1948). Sur les operations definies pour nombres reels. *Bulletin de la Societe Mathematique de France*, 76, 59-64.
[2] J. Aczel (1966). *Lectures on functional equations and their applications*. Academic Press.
[3] J. A. Baker (1980). The stability of the cosine equation. *Proceedings of the American Mathematical Society*, 80, 411-416.
[4] M. G. Bowler (1982). *Statistical mechanics*. Pergamon Press.
[5] S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci, and A. Rustichini (2020). Multinomial Logit processes and preference discovery: inside and outside the black box. Manuscript.
[6] R. Eisberg and R. Resnick (1985). *Quantum physics*. Wiley.
[7] E. T. Jaynes (1957). Information theory and statistical mechanics. *Physical Review*, 106, 620-630.
[8] K. Kanazawa, T. Sueshige, H. Takayasu, and M. Takayasu (2018). Derivation of the Boltzmann equation for financial Brownian motion: Direct observation of the collective motion of high-frequency traders. *Physical Review Letters*, 120, 138301.
[9] L. D. Landau and E. M. Lifshitz (1980). *Statistical physics*. Pergamon Press.
[10] E. L. Lehmann and J. P. Romano (2006). *Testing statistical hypotheses*. Springer.
[11] R. D. Luce (1959). *Individual choice behavior: a theoretical analysis*. Wiley.
[12] F. Matejka and A. McKay (2015). Rational inattention to discrete choices: a new foundation for the Multinomial Logit model. *American Economic Review*, 105, 272-298.
[13] A. Renyi (1955). On a new axiomatic theory of probability. *Acta Mathematica Hungarica*, 6, 285-335.
[14] K. Saito (2018). Axiomatizations of the Mixed Logit model. Manuscript.
[15] K. Sharp and F. Matschinsky (2015). Translation of Ludwig Boltzmann's paper "On the relationship between the second fundamental theorem of the mechanical theory of heat and probability calculations regarding the conditions for thermal equilibrium" Sitzungberichte der Kaiserlichen Akademie der Wissenschaften. Mathematisch-Naturwissen Classe. Abt. II, LXXVI 1877, pp. 373-435 (Wien. Ber. 1877, 76: 373-435). Reprinted in Wiss. Abhandlungen, Vol. II, reprint 42, p. 164-223, Barth, Leipzig, 1909. *Entropy*, 17, 1971-2009.
[16] K. E. Train (2009). *Discrete choice methods with simulation*. Cambridge University Press.
[17] C. Tsallis (1988). Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, 52, 479-487.
[18] V. M. Yakovenko and J. B. Rosser Jr (2009). Statistical mechanics of money, wealth, and income. *Reviews of Modern Physics*, 81, 1703-1725.

Appendix A. Proofs of the main results

This appendix presents the proofs of Theorem 1, Theorem 2, and Proposition 3.

A theorem of Aczel [1] characterizes continuous concatenations.

Theorem 6 (Aczel). *A binary operation $\oplus$ on $\mathbb{R}_+$ is a continuous concatenation if and only if there exists an increasing bijection $f : \mathbb{R}_+ \to \mathbb{R}_+$ such that*

$$t \oplus s = f^{-1}\left(f\left(t\right) + f\left(s\right)\right) \qquad \forall t, s \in \mathbb{R}_+$$

In this case, $f\left(0\right) = 0$ and f is strictly increasing and continuous.

The function f is said be a generator for $\oplus$, which is then denoted by $\oplus_f$.

Lemma 7. *If $p : (0, \infty) \times X \times \mathcal{A} \to \mathbb{R}_+$ is a random state function that satisfies A.1, A.3, A.4, and A.5, then:*

(i) *the relation defined on X by $a \succsim b$ if and only if $p_0\left(a, b\right) > 0$ is such that*

$$\begin{aligned} a \succ b &\iff p_0\left(a, b\right) > p_0\left(b, a\right) \\ &\iff p_0\left(a, b\right) = 1 \text{ and } p_0\left(b, a\right) = 0 \\ a \sim b &\iff p_0\left(a, b\right) = p_0\left(b, a\right) \\ &\iff p_0\left(a, b\right) = p_0\left(b, a\right) \in \{1, 1/2\} \\ b \succ a &\iff p_0\left(a, b\right) < p_0\left(b, a\right) \\ &\iff p_0\left(a, b\right) = 0 \text{ and } p_0\left(b, a\right) = 1 \end{aligned}$$

(ii) *given any $a, b \in X$, the function $\varphi_{a,b} : (0, \infty) \to (0, \infty)$ defined by*

$$\varphi_{a,b}\left(t\right) = r_{1/t}\left(a, b\right) \qquad \forall t \in (0, \infty)$$

is continuous and either diverges to ∞ as $t \to \infty$ (if $a \succ b$) or is constantly equal to 1 (if $a \sim b$) or vanishes as $t \to \infty$ (if $b \succ a$).

Proof A.1 and A.3 imply that $p_0\left(\cdot \mid \{a, b\}\right)$ is a probability distribution (supported) on $\{a, b\}$, for all $a, b \in X$. The proof is made pedantic by the fact that, if $a = b$, then $\{a, b\} = \{a\} = \{b\}$ and

$$p_0\left(a, b\right) + p_0\left(b, a\right) = p_0\left(a \mid \{a\}\right) + p_0\left(b \mid \{b\}\right) = 2$$

else $a \neq b$ and

$$p_0\left(a, b\right) + p_0\left(b, a\right) = 1$$

(i) By definition, $a \succ b$ iff $a \succsim b$ and not $b \succsim a$, that is, $p_0\left(a, b\right) > 0$ and $p_0\left(b, a\right) \leq 0$.

- Assume $a \succ b$, then $p_0\left(a, b\right) > 0$ and $p_0\left(b, a\right) \leq 0$ imply $p_0\left(a, b\right) > p_0\left(b, a\right)$.
- Assume $p_0\left(a, b\right) > p_0\left(b, a\right)$. This is impossible if $a = b$, therefore $a \neq b$ and $p_0\left(a, b\right) > 0$. If it held $p_0\left(b, a\right) > 0$, then $p_0\left(a, b\right) + p_0\left(b, a\right) = 1$ would imply $p_0\left(a, b\right), p_0\left(b, a\right) \in (0, 1)$, and A.5 would yield $p_0\left(a, b\right) = 1/2 = p_0\left(b, a\right)$, a contradiction. Then it must be $p_0\left(b, a\right) = 0$ and $p_0\left(a, b\right) = p_0\left(a, b\right) + p_0\left(b, a\right) = 1$.
- Assume $p_0\left(a, b\right) = 1$ and $p_0\left(b, a\right) = 0$, then $p_0\left(a, b\right) = 1$ and $p_0\left(b, a\right) \leq 0$, and $a \succ b$.

By definition, $a \sim b$ iff $a \succsim b$ and also $b \succsim a$, that is, $p_0\left(a, b\right) > 0$ and $p_0\left(b, a\right) > 0$.

- Assume $a \sim b$. If $a = b$, then $p_0(a,b) = 1 = p_0(b,a)$. Else $a \neq b$, $p_0(a,b), p_0(b,a) > 0$, and $p_0(a,b) + p_0(b,a) = 1$, then $p_0(a,b), p_0(b,a) \in (0,1)$, and A.5 yields $p_0(a,b) = 1/2 = p_0(b,a)$.
- Assume $p_0(a,b) = p_0(b,a)$. If $a = b$, then $p_0(a,b) = p_0(b,a) = 1$. Else $a \neq b$, and $p_0(a,b) + p_0(b,a) = 1$, then $2p_0(a,b) = 1$ and $2p_0(b,a) = 1$, that is, $p_0(a,b) = p_0(b,a) = 1/2$.
- Assume $p_0(a,b) = p_0(b,a) \in \{1, 1/2\}$, then $p_0(a,b), p_0(b,a) > 0$, and $a \sim b$.

The case $b \succ a$ follows from the case $a \succ b$ exchanging the roles of the states.

(ii) Given any $t \in (0,\infty)$, $\varphi_{a,b}(t) = r_{1/t}(a,b) = p_{1/t}(a,b)/p_{1/t}(b,a) \in (0,\infty)$ for all $a, b \in X$ because $p_{1/t}(\cdot \mid \{a,b\})$ is a positive probability distribution on $\{a,b\}$, thus $\varphi_{a,b} : (0,\infty) \to (0,\infty)$ is well defined. Moreover, by A.3, $\varphi_{a,b}$ is also continuous on $(0,\infty)$.

- If $a \succ b$, then $p_0(a,b) = 1$ and $p_0(b,a) = 0$, so $a \neq b$ and
$$\lim_{t\to\infty} \varphi_{a,b}(t) = \lim_{t\to\infty} \frac{p_{1/t}(a,b)}{p_{1/t}(b,a)} = \lim_{t\to\infty} \frac{1 - p_{1/t}(b,a)}{p_{1/t}(b,a)} = \infty$$
hence $\varphi_{a,b}$ diverges at ∞ as $t \to \infty$.

For later reference, note that so far A.4 has not been used.

- If $a \sim b$, and *per contra* $\varphi_{a,b}(t) \neq 1$ for some $t \in (0,\infty)$, then
  - either $\varphi_{a,b}(t) > 1$, thus $p_{1/t}(a,b) > p_{1/t}(b,a)$ and, by A.4, $p_0(a,b) > p_0(b,a)$, contradicting $a \sim b$,
  - or $\varphi_{a,b}(t) < 1$, thus $p_{1/t}(a,b) < p_{1/t}(b,a)$ and, by A.4, $p_0(a,b) < p_0(b,a)$, contradicting $a \sim b$,

  in conclusion, $\varphi_{a,b}(t) = 1$ for all $t \in (0,\infty)$.
- If $b \succ a$, the thesis follows because $\varphi_{a,b} = 1/\varphi_{b,a}$. ■

Proof of Theorem 2 Let p be a random state function that satisfies A.1–A.5 and A.7. As in Lemma 7, define, for all $a, b \in X$,
$$\varphi_{a,b}(t) = r_{1/t}(a,b) \qquad \forall t \in (0,\infty)$$
Also let $f : [0,\infty) \to [0,\infty)$ be a generator of a continuous concatenation $\oplus = \oplus_f$ for which A.7 holds. Set $g = f^{-1}$. By Theorem 6, $g : [0,\infty) \to [0,\infty)$ is a continuous and strictly increasing bijection such that $g(0) = 0$.

Next we show that, given any $a, b \in X$,
$$\varphi_{a,b}(g(t+s)) = \varphi_{a,b}(g(t))\,\varphi_{a,b}(g(s)) \qquad \forall t, s \in (0,\infty) \tag{12}$$
Three cases have to be considered, depending on whether $a \succ b$, $a \sim b$, or $b \succ a$ according to the relation $\succsim$ defined in Lemma 7.

- If $a \succ b$, then $\varphi_{a,b}$ is unbounded above and so is $\varphi_{a,b} \circ g : (0,\infty) \to (0,\infty)$. Moreover, by A.7, there exists $M > 0$ such that, for all $t, s \in (0,\infty)$
$$\left| r_{\frac{1}{t\oplus s}}(a,b) - r_{\frac{1}{t}}(a,b)\, r_{\frac{1}{s}}(a,b) \right| < M$$
$$\left| r_{\frac{1}{g(g^{-1}(t)+g^{-1}(s))}}(a,b) - r_{\frac{1}{t}}(a,b)\, r_{\frac{1}{s}}(a,b) \right| < M$$

hence, for all $t', s' \in (0, \infty)$, choosing $t = g(t')$ and $s = g(s')$, we have

$$\left| r_{\frac{1}{g(g^{-1}(g(t'))+g^{-1}(g(s')))}}(a,b) - r_{\frac{1}{g(t')}}(a,b)\, r_{\frac{1}{g(s')}}(a,b) \right| < M$$

$$\left| r_{\frac{1}{g(t'+s')}}(a,b) - r_{\frac{1}{g(t')}}(a,b)\, r_{\frac{1}{g(s')}}(a,b) \right| < M$$

$$\left| \varphi_{a,b}\left(g\left(t'+s'\right)\right) - \varphi_{a,b}\left(g\left(t'\right)\right) \varphi_{a,b}\left(g\left(s'\right)\right) \right| < M$$

But $(0, \infty)$ is a semigroup with respect to usual addition and $\varphi_{a,b} \circ g$ is unbounded above. Therefore, Theorem 1 of Baker [3] implies that (12) holds.

- If $a \sim b$, then $\varphi_{a,b}(t) = 1$ for all $t \in (0, \infty)$ and (12) holds.
- Else, $b \succ a$ and, as the first point shows,

$$\varphi_{b,a}(g(t+s)) = \varphi_{b,a}(g(t))\, \varphi_{b,a}(g(s))$$

for all $t, s \in (0, \infty)$, but then

$$\begin{aligned}\varphi_{a,b}(g(t+s)) &= \frac{1}{\varphi_{b,a}(g(t+s))} \\ &= \frac{1}{\varphi_{b,a}(g(t))\, \varphi_{b,a}(g(s))} \\ &= \varphi_{a,b}(g(t))\, \varphi_{a,b}(g(s))\end{aligned}$$

and (12) holds again.

Summing up, the functional equation (12) holds for all $a, b \in X$. Continuity of $\varphi_{a,b} \circ g$, its strict positivity, and (12), imply that

$$\varphi_{a,b}(g(t)) = e^{v(a,b)t} \qquad \forall t \in (0, \infty)$$

for a unique $v(a,b) \in \mathbb{R}$ (see, e.g., Theorem 2.1.2.1 of Aczel [2]). It follows that $\varphi_{a,b}(s) = \varphi_{a,b}(g(f(s))) = e^{v(a,b)f(s)}$ for all $s \in (0, \infty)$.

Now fix some $a^* \in X$ and define $E : X \to \mathbb{R}$ by $E(x) = -v(x, a^*)$ for all $x \in X$. Given any $t \in (0, \infty)$ and any $x, y \in X$, by A.1, A.2, and Theorem 2 of Luce [11], it follows that

$$\begin{aligned}\varphi_{x,y}(t) &= r_{1/t}(x,y) = r_{1/t}(x,a^*)\, r_{1/t}(a^*,y) = \frac{r_{1/t}(x,a^*)}{r_{1/t}(y,a^*)} \\ &= \frac{\varphi_{x,a^*}(t)}{\varphi_{y,a^*}(t)} = \frac{e^{v(x,a^*)f(t)}}{e^{v(y,a^*)f(t)}} = \frac{e^{-E(x)f(t)}}{e^{-E(y)f(t)}}\end{aligned}$$

By Theorem 3 of Luce [11], for every $t \in (0, \infty)$, $A \in \mathcal{A}$, and $a \in A$, arbitrarily choosing $c^* \in A$,

$$\begin{aligned}p_t(a \mid A) &= \frac{r_t(a,c^*)}{\sum_{b \in A} r_t(b,c^*)} = \frac{\varphi_{a,c^*}(1/t)}{\sum_{b \in A} \varphi_{b,c^*}(1/t)} \\ &= \frac{\frac{e^{-E(a)f(1/t)}}{e^{-E(c^*)f(1/t)}}}{\sum_{b \in A} \frac{e^{-E(b)f(1/t)}}{e^{-E(c^*)f(1/t)}}} = \frac{e^{-f\left(\frac{1}{t}\right)E(a)}}{\sum_{b \in A} e^{-f\left(\frac{1}{t}\right)E(b)}}\end{aligned}$$

and (2) holds for $\kappa(t) = 1/f(1/t)$ (because $p_t(a \mid A) = 0$ for $a \notin A$ by A.1).

NB 1 So far we have shown that: If p is random state function that satisfies A.1–A.5 and A.7 (with respect to $\oplus_f$); then, setting $\kappa(t) = 1/f(1/t)$ for all $t \in (0, \infty)$, there exists $E : X \to \mathbb{R}$ such that

$$p_t(a \mid A) = \frac{e^{-\frac{E(a)}{\kappa(t)}}}{\sum_{b \in A} e^{-\frac{E(b)}{\kappa(t)}}} \delta_a(A)$$

for all $(t, a, A) \in (0, \infty) \times X \times \mathcal{A}$.
Moreover, since $f_{|(0,\infty)}$ is a continuous and strictly increasing bijection from $(0, \infty)$ to $(0, \infty)$, and $s \mapsto 1/s$ is a continuous and strictly decreasing bijection from $(0, \infty)$ to $(0, \infty)$, then $\kappa : t \mapsto 1/f(1/t)$ a continuous and strictly increasing bijection from $(0, \infty)$ to $(0, \infty)$.

This proves the "only if" part of the statement.

As to the "if" part, assume that (2) holds. It is routine to check that p satisfies A.1–A.5. To prove that also A.7 holds, define $\phi : [0, \infty) \to [0, \infty)$ by setting $\phi(v) = 1/\kappa(1/v)$ for all $v > 0$, and $\phi(0) = 0$. Since k is an increasing bijection from $(0, \infty)$ to $(0, \infty)$, so is $\phi_{|(0,\infty)}$. But then $\phi : [0, \infty) \to [0, \infty)$ is an increasing bijection too. Then

$$t \oplus s = \phi^{-1}[\phi(t) + \phi(s)] \qquad \forall t, s \in [0, \infty) \tag{13}$$

is a (well defined) binary operation on $\mathbb{R}_+$. Theorem 6 guarantees that $\oplus = \oplus_\phi$ is indeed a continuous concatenation. With this, given $a, b \in X$, for all $t, s \in (0, \infty)$

$$r_{\frac{1}{t \oplus s}}(a, b) = e^{-\frac{1}{\kappa\left(\frac{1}{t \oplus s}\right)}[E(a) - E(b)]}$$

but, by (13), $t \oplus s > 0$, hence, by definition of ϕ,

$$\begin{aligned} r_{\frac{1}{t \oplus s}}(a, b) &= e^{-\phi(t \oplus s)[E(a) - E(b)]} = e^{-(\phi(t) + \phi(s))[E(a) - E(b)]} \\ &= e^{-\phi(t)[E(a) - E(b)]} e^{-\phi(s)[E(a) - E(b)]} \\ &= e^{-\frac{1}{\kappa(1/t)}[E(a) - E(b)]} e^{-\frac{1}{\kappa(1/s)}[E(a) - E(b)]} \\ &= r_{1/t}(a, b)\, r_{1/s}(a, b) \end{aligned}$$

A fortiori, A.7 holds, with respect to $\oplus_\phi$, where $\phi(v) = 1/\kappa(1/v)$ for all $v > 0$, and $\phi(0) = 0$. Actually, we proved a stronger fact:

NB 2 Given a function $E : X \to \mathbb{R}$ and an increasing bijection $\kappa : (0, \infty) \to (0, \infty)$, the function defined by

$$p_t(a \mid A) = \frac{e^{-\frac{E(a)}{\kappa(t)}}}{\sum_{b \in A} e^{-\frac{E(b)}{\kappa(t)}}} \delta_a(A)$$

for all $(t, a, A) \in (0, \infty) \times X \times \mathcal{A}$ is a random state function that satisfies A.1–A.5 and A.7 (with respect to $\oplus_\phi$, where $\phi(v) = 1/\kappa(1/v)$ for all $v > 0$, and $\phi(0) = 0$).

This concludes the proof of the first part of the statement.

Now assume that (2) holds for a function $E : X \to \mathbb{R}$ and an increasing bijection $\kappa : (0,\infty) \to (0,\infty)$. Note that

$$r_t(a,b) = \exp\left(-\frac{1}{\kappa(t)}[E(a)-E(b)]\right)$$
$$\frac{1}{\kappa(t)}[E(a)-E(b)] = -\ln r_t(a,b) = \ln r_t(b,a)$$

for all $(t,a,b) \in (0,\infty) \times X^2$.

If p is uniform, then $\ln r_t(b,a) = 0$ for all $(t,a,b) \in (0,\infty) \times X^2$, and strict positivity of κ implies E is constant. The converse follows immediately from (2).

Else, E is not constant. Let $\tilde{E} : X \to \mathbb{R}$ and $\tilde{\kappa} : (0,\infty) \to (0,\infty)$ also represent p as in (2), then

$$\frac{1}{\kappa(t)}[E(a)-E(b)] = \ln r_t(b,a) = \frac{1}{\tilde{\kappa}(t)}\left[\tilde{E}(a)-\tilde{E}(b)\right]$$

for all $(t,a,b) \in (0,\infty) \times X^2$. Arbitrarily choose $(t^*,a^*,b^*) \in (0,\infty) \times X^2$ such that $E(a^*) > E(b^*)$. Then:

(i) For all $a \in A$,

$$\frac{1}{\kappa(t^*)}[E(a)-E(b^*)] = \ln r_{t^*}(b^*,a) = \frac{1}{\tilde{\kappa}(t^*)}\left[\tilde{E}(a)-\tilde{E}(b^*)\right]$$

hence

$$\tilde{E}(a) = \underbrace{\frac{\tilde{\kappa}(t^*)}{\kappa(t^*)}}_{m^*}E(a) + \underbrace{\tilde{E}(b^*) - \frac{\tilde{\kappa}(t^*)}{\kappa(t^*)}E(b^*)}_{q^*}$$

and, for all $t \in (0,\infty)$,

$$\frac{1}{\kappa(t)}[E(a^*)-E(b^*)] = \ln r_t(b^*,a^*) = \frac{1}{\tilde{\kappa}(t)}\left[\tilde{E}(a^*)-\tilde{E}(b^*)\right]$$
$$\frac{1}{\kappa(t)}[E(a^*)-E(b^*)] = \frac{1}{\tilde{\kappa}(t)}[m^*E(a^*)-m^*E(b^*)]$$
$$\tilde{\kappa}(t) = m^*\kappa(t)$$

thus there exist $m > 0$ and $q \in \mathbb{R}$ such that $\tilde{E} = mE + q$ and $\tilde{\kappa} = m\kappa$. This proves the "only if" part of point (i). The "if" part is trivial.

(ii) By NB 2, under (2), the binary operation defined by

$$t \oplus_\phi s = \phi^{-1}[\phi(t)+\phi(s)] \qquad \forall t,s \in [0,\infty)$$

where $\phi : [0,\infty) \to [0,\infty)$ is given by $\phi(v) = 1/\kappa(1/v)$ for all $v > 0$ and $\phi(0) = 0$, is a concatenation for which A.7 holds. By NB 1, if $\oplus = \oplus_f$ is a concatenation for which A.7

holds, then setting $\tilde{\kappa}(t) = 1/f(1/t)$ for all $t \in (0,\infty)$, there exists $\tilde{E} : X \to \mathbb{R}$ such that

$$p_t(a \mid A) = \frac{e^{-\frac{\tilde{E}(a)}{\tilde{\kappa}(t)}}}{\sum_{b\in A} e^{-\frac{\tilde{E}(b)}{\tilde{\kappa}(t)}}}\delta_a(A)$$

for all $(t,a,A) \in (0,\infty) \times X \times \mathcal{A}$. By point (i), there exist $m > 0$ and $q \in \mathbb{R}$ such that $\tilde{E} = mE + q$ and $\tilde{\kappa} = m\kappa$; therefore, for all $t \in (0,\infty)$,

$$\frac{1}{f(1/t)} = \tilde{\kappa}(t) = m\kappa(t) = \frac{m}{\phi(1/t)}$$

hence $f = \phi/m$ on $(0,\infty)$, and $f(0) = 0 = \phi(0)/m$ by Theorem 6. Finally, $f = \phi/m$ implies $\oplus_f = \oplus_\phi$, concluding the proof of (ii). ■

Proof of Theorem 1 If p is a random state function that satisfies A.1–A.5 and A.6, then it also satisfies A.7 with respect to $\oplus_f$ where $f(t) = t/k$ and k is the Boltzmann constant. By NB 1 of the previous proof, setting $\kappa(t) = 1/f(1/t)$ for all $t \in (0,\infty)$, it follows $\kappa(t) = kt$ and there exists $E : X \to \mathbb{R}$ such that

$$p_t(a \mid A) = \frac{e^{-\frac{E(a)}{kt}}}{\sum_{b\in A} e^{-\frac{E(b)}{kt}}}\delta_a(A)$$

for all $(t,a,A) \in (0,\infty) \times X \times \mathcal{A}$. The converse is routine.

As to uniqueness of the representation, by point (i) of Theorem 2, if $\tilde{E} : X \to \mathbb{R}$, and $\tilde{\kappa}(t) = kt$, also represent p as in (1), then there exist $m > 0$ and $q \in \mathbb{R}$ such that $\tilde{E} = mE + q$ and $\tilde{\kappa} = m\kappa$, but this means $kt = mkt$ for all $t > 0$, that is $m = 1$. Again, the converse is routine. ■

Proof of Proposition 3 *(ii) is equivalent to (i).* There exists $t \in (0,\infty)$ such that (3) holds if and only if

$$\begin{aligned}
&\exists t \in (0,\infty) : p_{\alpha t}(\alpha a + (1-\alpha)b, b) \geq p_t(a,b)\\
\iff\ &\exists t \in (0,\infty) : r_{\alpha t}(\alpha a + (1-\alpha)b, b) \geq r_t(a,b)\\
\iff\ &\exists t \in (0,\infty) : r_{\alpha t}(b, \alpha a + (1-\alpha)b) \leq r_t(b,a)\\
\iff\ &\exists t \in (0,\infty) : e^{\frac{1}{k\alpha t}[E(\alpha a+(1-\alpha)b)-E(b)]} \leq e^{\frac{1}{kt}[E(a)-E(b)]}\\
\iff\ &E(\alpha a + (1-\alpha)b) \leq \alpha E(a) + (1-\alpha)E(b)
\end{aligned}$$

for all $(a,b,\alpha) \in X \times X \times (0,1)$.

(i) implies (iii). Given any $s \in (0,\infty)$, $A \in \mathcal{A}$, $b \in A$, and $\eta > 1$,

$$p_s\left(b \;\middle|\; \frac{1}{\eta}A + \left(1 - \frac{1}{\eta}\right)b\right) = \frac{1}{\sum_{a\in A} e^{-\frac{1}{ks}\left[E\left(\frac{1}{\eta}a+\left(1-\frac{1}{\eta}\right)b\right)-E(b)\right]}}$$

but convexity of E implies $E((1/\eta)a + (1-(1/\eta))b) - E(b) \leq (1/\eta)(E(a) - E(b))$ hence

$$-\frac{1}{ks}\left[E((1/\eta)a + (1-(1/\eta))b) - E(b)\right] \geq -\frac{1}{k\eta s}\left[E(a) - E(b)\right]$$

for all $a \in A$, and $p_s(b \mid (1/\eta)A + (1-(1/\eta))b) \leq p_{\eta s}(b \mid A)$.

(iii) implies (iv). Trivial.

(iv) implies (i). To prove convexity, it is sufficient to check that, given any $\alpha \in (0,1)$,

$$E(\alpha x + (1-\alpha) y) \leq \alpha E(x) + (1-\alpha) E(y) \tag{14}$$

for all $x, y \in X$ such that $E(y) \geq E(x)$.[9] Now, arbitrarily choose $s \in (0,\infty)$. If $E(y) \geq E(x)$, then $y \in \arg\min_{a \in \{x,y\}} p_{s/\alpha}(a \mid \{x,y\})$, then (5), with $\eta = 1/\alpha$, yields

$$\begin{aligned} & p_s(y \mid \alpha\{x,y\} + (1-\alpha) y) \leq p_{s/\alpha}(y \mid \{x,y\}) \\ \implies & r_s(y, \alpha x + (1-\alpha) y) \leq r_{s/\alpha}(y, x) \\ \implies & \frac{1}{ks}[E(\alpha x + (1-\alpha) y) - E(y)] \leq \frac{\alpha}{ks}[E(x) - E(y)] \end{aligned}$$

for all $\alpha \in (0,1)$, which implies (14)). ■

Appendix B. Additional axiomatizations

This appendix presents the proofs of Theorem 4 and Corollary 5.

The proof of Theorem 4 is quite involved. We start with some preliminary results of independent interest.

B.1. **Identification of the energy and noise functions in Theorem 2.**

First observe that when p is *uniform* the axioms of Theorem 2 hold, E is constant, and κ is undetermined. Otherwise, we have the following result.

Proposition 8. *Let p be a random state function that satisfies $p_{\bar{v}}(\bar{c}, \bar{d}) > p_{\bar{v}}(\bar{d}, \bar{c})$ for some $\bar{v} \in (0,\infty)$ and $\bar{c}, \bar{d} \in X$. If A.1–A.5 are not violated, then A.7 is satisfied if and only if representation (2) holds with*

$$\tilde{E}(a) = \ln r_{\bar{v}}(\bar{c}, a) \qquad \text{and} \qquad \tilde{\kappa}(t) = \frac{\ln r_{\bar{v}}(\bar{c}, \bar{d})}{\ln r_t(\bar{c}, \bar{d})}$$

for all $(t, a) \in (0,\infty) \times X$.

This result shows how the energy and noise functions can be directly derived from the random state function (provided of course A.1–A.5, and A.7 are not violated).

Proof of Proposition 8 If A.7 is satisfied, by Theorem 2 there exist a function $E : X \to \mathbb{R}$ and an increasing bijection $\kappa : (0,\infty) \to (0,\infty)$ such that p is represented by (2). Moreover, $p_{\bar{v}}(\bar{c}, \bar{d}) > p_{\bar{v}}(\bar{d}, \bar{c})$ implies $E(\bar{d}) > E(\bar{c})$.

For all $a \in A$,

$$r_{\bar{v}}(a, \bar{c}) = \exp\left(-\frac{1}{\kappa(\bar{v})}[E(a) - E(\bar{c})]\right)$$

$$\frac{1}{\kappa(\bar{v})}[E(a) - E(\bar{c})] = -\ln r_{\bar{v}}(a, \bar{c}) = \ln r_{\bar{v}}(\bar{c}, a)$$

[9] In fact, if $E(\bar{x}) > E(\bar{y})$, (14) yields, for any $\beta \in (0,1)$, $E(\beta\bar{y} + (1-\beta)\bar{x}) \leq \beta E(\bar{y}) + (1-\beta) E(\bar{x})$.

hence $\ln r_{\bar{v}}(\bar{c}, a) = mE(a) + q$, with $m = 1/\kappa(\bar{v})$ and $q = -E(\bar{c})/\kappa(\bar{v})$. For all $t \in (0, \infty)$,

$$\frac{\ln r_{\bar{v}}\left(\bar{c}, \bar{d}\right)}{\ln r_t\left(\bar{c}, \bar{d}\right)} = \frac{-\frac{1}{\kappa(\bar{v})}\left[E(\bar{c}) - E\left(\bar{d}\right)\right]}{-\frac{1}{\kappa(t)}\left[E(\bar{c}) - E\left(\bar{d}\right)\right]} = \frac{1}{\kappa(\bar{v})}\kappa(t) = m\kappa(t)$$

Point (i) of Theorem 2 implies that (2) holds, with $\tilde{E}(\cdot) = mE(\cdot) + q = \ln r_{\bar{v}}(\bar{c}, \cdot)$ and $\tilde{\kappa}(\cdot) = m\kappa(\cdot) = \ln r_{\bar{v}}\left(\bar{c}, \bar{d}\right) / \ln r_{\cdot}\left(\bar{c}, \bar{d}\right)$.

The converse follows from Theorem 2 too: if representation (2) holds,[10] then A.7 is satisfied. ∎

B.2. **Testability of Axioms A.4 and A.7.** Axioms A.4 and A.7 are difficult to test: The first requires observation of the system at the limit temperature 0 for all $a, b \in X$. The second contains an existential quantifier. Axioms A.8 and A.9 do not present these drawbacks.

The next result permits to replace A.4 and A.7 in Theorem 2 with A.8 and A.9 used in Theorem 4.

Proposition 9. *Let $p : (0, \infty) \times X \times \mathcal{A} \to \mathbb{R}_+$ be a random state function that satisfies A.1–A.3, and A.5. Then, p satisfies A.4 and A.7 if and only if it satisfies A.8 and A.9.*

The next Lemma uses the notation of Lemma 7.

Lemma 10. *If $p : (0, \infty) \times X \times \mathcal{A} \to \mathbb{R}_+$ is a random state function that satisfies A.1, A.3, A.5, and A.8, then, given any $a, b \in X$:*

(i) *$a \succ b$ if and only if $\varphi_{a,b}$ is an increasing bijection from $(0, \infty)$ to $(1, \infty)$;*
(ii) *$a \sim b$ if and only if $\varphi_{a,b}$ is constantly equal to 1;*
(iii) *$a \prec b$ if and only if $\varphi_{a,b}$ is a decreasing bijection from $(0, \infty)$ to $(0, 1)$.*

In particular, all the above monotonicity and bijectivity properties are maintained when $\varphi_{a,b}$ is extended to $[0, \infty)$ by setting $\varphi_{a,b}(0) = 1$.

Proof By the arguments adopted in the proof of Lemma 7, we have that, given any $a, b \in X$, the function

$$\begin{array}{rccc} \varphi_{a,b}: & (0,\infty) & \to & (0,\infty) \\ & t & \mapsto & r_{1/t}(a,b) \end{array}$$

is well defined, and continuous.

Fact 1. *If $r_\tau(a, b) > 1$ for some $\tau \in (0, \infty)$, then*

$$\begin{array}{rccc} r(a,b): & (0,\infty) & \to & (0,\infty) \\ & t & \mapsto & r_t(a,b) \end{array}$$

is strictly decreasing and everywhere strictly greater than 1, that is, $\varphi_{a,b}$ is strictly increasing and everywhere strictly greater than 1.

If $r_\tau(a, b) < 1$ for some $\tau \in (0, \infty)$, then

$$\begin{array}{rccc} r(a,b): & (0,\infty) & \to & (0,\infty) \\ & t & \mapsto & r_t(a,b) \end{array}$$

[10] At the risk of being pedantic, the sentence "representation (2) holds for some $\tilde{E}$ and $\tilde{\kappa}$" means that $\tilde{E} : X \to \mathbb{R}$ is a function, $\tilde{\kappa} : (0, \infty) \to (0, \infty)$ is an increasing bijection, and equation (2) holds for all $(t, a, A) \in (0, \infty) \times X \times \mathcal{A}$.

is strictly increasing and everywhere strictly smaller than 1*, that is,* $\varphi_{a,b}$ *is strictly decreasing and everywhere strictly smaller than* 1*.*

Proof Let $r_\tau(a,b) > 1$. If $r_t(a,b) \leq 1$ for some $t > \tau$, by A.8 it would follow $r_\tau(a,b) \leq r_t(a,b) \leq 1$, a contradiction. Then $r_t(a,b) > 1$, for all $t \in [\tau,\infty)$. Now, given any $s \in (0,\infty)$, taking $t \in [\tau,\infty)$ such that $s < t$, by A.8 it follows $r_s(a,b) > r_t(a,b) > 1$. Therefore, $r_t(a,b) > 1$, for all $t \in (0,\infty)$. But then, given any $s < t$ in $(0,\infty)$, since $r_t(a,b) > 1$, by A.8 it follows $r_s(a,b) > r_t(a,b)$, and

$$\begin{array}{rccc} r(a,b): & (0,\infty) & \to & (0,\infty) \\ & t & \mapsto & r_t(a,b) \end{array}$$

is strictly decreasing, then $\varphi_{a,b}$ is strictly increasing.

Let $r_\tau(a,b) < 1$, then

$$r_\tau(b,a) = \frac{1}{r_\tau(a,b)} > 1$$

hence $r(b,a)$ is strictly decreasing, $r(a,b)$ strictly increasing, $\varphi_{a,b}$ strictly decreasing. □

(i) If $a \succ b$, again by arguments of the proof of Lemma 7, it follows that

$$\lim_{t\to\infty} r_{1/t}(a,b) = \lim_{t\to\infty} \varphi_{a,b}(t) = \infty$$

then $r_\tau(a,b) > 1$ for some $\tau \in (0,\infty)$, and $\varphi_{a,b}$ is strictly increasing. Finally, by A.8,

$$\lim_{t\to 0} \varphi_{a,b}(t) = \lim_{t\to\infty} r_t(a,b) = 1$$

and so $\varphi_{a,b}$ is an increasing bijection from $(0,\infty)$ to $(1,\infty)$.

Conversely, if $\varphi_{a,b}$ is an increasing bijection from $(0,\infty)$ to $(1,\infty)$, then

$$r_0(a,b) = \lim_{t\to 0} r_t(a,b) = \lim_{t\to\infty} r_{1/t}(a,b) = \lim_{t\to\infty} \varphi_{a,b}(t) = \infty$$

But then it must be the case that $a \neq b$, and the above limit corresponds to

$$\lim_{t\to 0} \frac{1 - p_t(b,a)}{p_t(b,a)} = \infty$$

thus $p_0(b,a) = 0$ and $p_0(a,b) = 1$. Then, by definition of $\succsim$, $a \succ b$.

(ii) If $a \sim b$ and $a = b$, then obviously, $\varphi_{a,b}(t) = p_{1/t}(a,b)/p_{1/t}(b,a) = 1$, irrespective of $t \in (0,\infty)$. Else if $a \sim b$ and $a \neq b$, by point (i) if Lemma 7 we have that $p_0(a,b) = p_0(b,a)$, and so

$$\lim_{t\to\infty} \varphi_{a,b}(t) = \lim_{t\to\infty} \frac{p_{1/t}(a,b)}{p_{1/t}(b,a)} = \lim_{t\to 0} \frac{p_t(a,b)}{p_t(b,a)} = \frac{p_0(a,b)}{p_0(b,a)} = 1 \tag{15}$$

If $\varphi_{a,b}(\bar{t}) > 1$ for some $\bar{t} \in (0,\infty)$, then $r_\tau(a,b) > 1$ for some $\tau \in (0,\infty)$ (say, $\tau = 1/\bar{t}$), then $\varphi_{a,b}$ is strictly increasing, which contradicts (15), because it implies $\lim_{t\to\infty} \varphi_{a,b}(t) \geq \varphi_{a,b}(\bar{t}) > 1$. If $\varphi_{a,b}(\bar{t}) < 1$ for some $\bar{t} \in (0,\infty)$, then $r_\tau(a,b) < 1$ for some $\tau \in (0,\infty)$ (say, $\tau = 1/\bar{t}$), then $\varphi_{a,b}$ is strictly decreasing, which contradicts (15). Therefore, $\varphi_{a,b}(t) = 1$, irrespective of $t \in (0,\infty)$.

Conversely, if $\varphi_{a,b}(t) \equiv 1$, then $r_t(a,b) \equiv 1$, hence $p_t(a,b) \equiv p_t(b,a)$, and $p_0(a,b) = p_0(b,a)$, thus $a \sim b$.

(iii) $a \prec b$ iff $b \succ a$ iff $\varphi_{b,a}$ is an increasing bijection from $(0,\infty)$ to $(1,\infty)$ iff $\varphi_{a,b} = 1/\varphi_{b,a}$ is a decreasing bijection from $(0,\infty)$ to $(0,1)$. ■

By the previous arguments, and since, by Lemma 7, $\succsim$ is a trichotomy, we have the following:

Corollary 11. *If $p : (0,\infty) \times X \times \mathcal{A} \to \mathbb{R}_+$ is a random state function that satisfies A.1, A.3, A.5, and A.8, then, given any $a, b \in X$:*

(i) *$a \succ b$ if and only if $r_t(a,b) > 1$ for some/all $t \in (0,\infty)$;*
(ii) *$a \sim b$ if and only if $r_t(a,b) = 1$ for some/all $t \in (0,\infty)$;*
(iii) *$a \prec b$ if and only if $r_t(a,b) < 1$ for some/all $t \in (0,\infty)$.*

Proof of Proposition 9 Assume p is not uniform (the uniform case is left to the reader). If p satisfies A.4 and A.7, then using the representation provided by Theorem 2, it is routine to show that it satisfies A.8 and A.9. We only prove the converse.

As to A.4, let $(t,a,b) \in (0,\infty) \times X^2$ be such that $p_t(a,b) > p_t(b,a)$. Then $a \neq b$ and $r_t(a,b) > 1$, by the previous results, $\varphi_{a,b}(t) = r_{1/t}(a,b)$ is an increasing bijection from $(0,\infty)$ to $(1,\infty)$, then

$$\lim_{s\to 0} \frac{p_s(a,b)}{1-p_s(a,b)} = \lim_{s\to 0} r_s(a,b) = \lim_{t\to\infty} \varphi_{a,b}(t) = \infty$$

thus $p_0(a,b) = 1 > 0 = p_0(b,a)$. As wanted.

As to A.7. Given any $a, b \in X$, set $\varphi_{a,b}(0) = 1$ as in Lemma 10. Denote $w_t(a,b) = \ln \varphi_{a,b}(t)$, for all $(t,a,b) \in [0,\infty) \times X^2$. Arbitrarily choose $\hat{a} \succ \hat{b} \in X$, so that $\varphi_{\hat{a},\hat{b}} : [0,\infty) \to [1,\infty)$ is an increasing bijection, and notice that the function

$$f(t) = \ln \varphi_{\hat{a},\hat{b}}(t) = w_t(\hat{a},\hat{b}) \qquad \forall t \in [0,\infty) \tag{16}$$

is an increasing bijection onto $[0,\infty)$, so $f(0) = 0$ and $f_{|(0,\infty)}$ is an increasing bijection onto $(0,\infty)$. The next steps verify that p satisfies A.7 with respect to $\oplus_f$.

Note that, given any $t, s \in (0,\infty)$, we have

$$w_{\underbrace{f^{-1}(f(t)+f(s))}_{\tau}}(\hat{a},\hat{b}) = f\left(\underbrace{f^{-1}(f(t)+f(s))}_{\tau}\right) \tag{17}$$

$$= f(t) + f(s) = w_t(\hat{a},\hat{b}) + w_s(\hat{a},\hat{b}) \tag{18}$$

Next we show that (17) and A.9 imply

$$w_{f^{-1}(f(t)+f(s))}(a,b) = w_t(a,b) + w_s(a,b) \tag{19}$$

for all $a, b \in X$ and all $t, s \in (0,\infty)$. Given any $c, d, x, y \in X$ and any $s, t, \tau \in (0,\infty)$ such that $w_\tau(c,d) > 0$ and $w_\tau(x,y) > 0$, we have $r_{1/\tau}(c,d) = e^{w_\tau(c,d)} > 1$ and $r_{1/\tau}(x,y) = e^{w_\tau(x,y)} > 1$, hence, by A.9,

$$\begin{aligned} & r_{1/\tau}(c,d) > r_{1/t}(c,d)\, r_{1/s}(c,d) \\ & \qquad \iff r_{1/\tau}(x,y) > r_{1/t}(x,y)\, r_{1/s}(x,y) \\ & w_t(c,d) > w_t(c,d) + w_s(c,d) \\ & \qquad \iff w_\tau(x,y) > w_t(x,y) + w_s(x,y) \end{aligned}$$

(the roles of (c,d) and (x,y) are symmetric in the axiom). By Corollary 11, if $w_{\hat{t}}(c,d) > 0$ and $w_{\hat{s}}(x,y) > 0$ for some $\hat{t}, \hat{s} \in (0,\infty)$, then $w_\tau(c,d) > 0$ and $w_\tau(x,y) > 0$ for all $\tau \in (0,\infty)$. Therefore, given any $c,d,x,y \in X$, if $w_{\hat{t}}(c,d) > 0$ and $w_{\hat{s}}(x,y) > 0$ for some $\hat{t}, \hat{s} \in (0,\infty)$, then, given any $s,t,\tau \in (0,\infty)$, it follows

$$w_\tau(c,d) \leq w_t(c,d) + w_s(c,d) \tag{20}$$

$$\iff w_\tau(x,y) \leq w_t(x,y) + w_s(x,y) \tag{21}$$

Moreover, as we argued for (16), since $c \succ d$ and $x \succ y$, the functions $h(t) = w_t(c,d)$ and $g(t) = w_t(x,y)$ are increasing bijections from $(0,\infty)$ to $(0,\infty)$ and (20) implies

$$\tau \leq h^{-1}(h(t) + h(s)) \iff \tau \leq g^{-1}(g(t) + g(s))$$

for all $s,t,\tau \in (0,\infty)$. But then, $h^{-1}(h(t) + h(s)) = g^{-1}(g(t) + g(s))$ for all $s,t \in (0,\infty)$. Hence, for all $s,t,\tau \in (0,\infty)$,

$$\tau = h^{-1}(h(t) + h(s)) \iff \tau = g^{-1}(g(t) + g(s))$$

that is, $h(\tau) = h(t) + h(s) \iff g(\tau) = g(t) + g(s)$.

Therefore:

- if $w_{\hat{t}}(c,d) > 0$ and $w_{\hat{s}}(x,y) > 0$ for some $\hat{t}, \hat{s} \in (0,\infty)$, then, given any $s,t,\tau \in (0,\infty)$, it holds

$$\begin{aligned} & w_\tau(c,d) = w_t(c,d) + w_s(c,d) \\ & \qquad \iff w_\tau(x,y) = w_t(x,y) + w_s(x,y) \end{aligned}$$

- if $w_{\hat{t}}(c,d) > 0$ and $w_{\hat{s}}(x,y) < 0$ for some $\hat{t}, \hat{s} \in (0,\infty)$, then, $w_{\hat{s}}(y,x) > 0$ and, given any $s,t,\tau \in (0,\infty)$, it holds

$$\begin{aligned} & w_\tau(c,d) = w_t(c,d) + w_s(c,d) \\ & \qquad \iff w_\tau(y,x) = w_t(y,x) + w_s(y,x) \\ & \qquad \iff -w_\tau(y,x) = -w_t(y,x) - w_s(y,x) \\ & \qquad \iff w_\tau(x,y) = w_t(x,y) + w_s(x,y) \end{aligned}$$

- if $w_{\hat{t}}(c,d) > 0$ and $w_{\hat{s}}(x,y) = 0$ for some $\hat{t}, \hat{s} \in (0,\infty)$, then, $\varphi_{x,y}$ is constantly equal to 1, and $w_\tau(x,y) = w_t(x,y) = w_s(x,y) = 0$, for all $s,t,\tau \in (0,\infty)$, thus, given any $s,t,\tau \in (0,\infty)$, it holds

$$\begin{aligned} & w_\tau(c,d) = w_t(c,d) + w_s(c,d) \\ & \qquad \Longrightarrow w_\tau(x,y) = w_t(x,y) + w_s(x,y) \end{aligned}$$

Summing up, since $\hat{a} \succ \hat{b}$, then, given any $s, t, \tau \in (0, \infty)$,

$$w_\tau(\hat{a}, \hat{b}) = w_t(\hat{a}, \hat{b}) + w_s(\hat{a}, \hat{b}) \tag{22}$$
$$\implies w_\tau(x, y) = w_t(x, y) + w_s(x, y) \tag{23}$$

for all $x, y \in X$. Now by (17)

$$w_{\underbrace{f^{-1}(f(t)+f(s))}_{\tau}}(\hat{a}, \hat{b}) = w_t(\hat{a}, \hat{b}) + w_s(\hat{a}, \hat{b}) \qquad \forall t, s \in (0, \infty)$$

and so (22) implies

$$w_{\underbrace{f^{-1}(f(t)+f(s))}_{\tau}}(x, y) = w_t(x, y) + w_s(x, y) \qquad \forall t, s \in (0, \infty)$$

and for all $x, y \in X$. Finally, for all $x, y \in X$ and all $t, s \in (0, \infty)$

$$\begin{aligned} r_{\frac{1}{t \oplus_f s}}(x, y) &= \varphi_{x,y}(t \oplus_f s) = \varphi_{x,y}\left(f^{-1}(f(t) + f(s))\right) \\ &= e^{w_{f^{-1}(f(t)+f(s))}(x,y)} = e^{w_t(x,y)} e^{w_s(x,y)} \\ &= \varphi_{x,y}(t)\, \varphi_{x,y}(s) = r_{\frac{1}{t}}(x, y)\, r_{\frac{1}{s}}(x, y) \end{aligned}$$

and A.7 holds. ■

B.3. **Proof of Theorem 4.** Theorem 2 and Proposition 9 yield the equivalence between the axioms and representation (6).

- If p is uniform, by Theorem 2, then E is constant, and κ is undetermined.
- Else p is not uniform, then it cannot be the case that $p_v(c, d) = p_v(d, c)$ for all $v \in (0, \infty)$ and all $c \neq d$ in X. Per contra, assume this is the case, then since A.2 holds, for all $A \in \mathcal{A}$, all $c, d \in A$, and all $v \in (0, \infty)$ it follows

  $$p_v(c \mid A) = p_v(c, d)\, p_v(\{c, d\} \mid A) = p_v(d, c)\, p_v(\{d, c\} \mid A) = p_v(d \mid A)$$

  yielding uniformity of p, a contradiction. Then, $p_{\bar{v}}\left(\bar{c}, \bar{d}\right) > p_{\bar{v}}\left(\bar{d}, \bar{c}\right)$ for some $\bar{v} \in (0, \infty)$ and $\bar{c}, \bar{d} \in X$, and so Proposition 8 delivers the explicit expressions (7) of the energy and noise functions. ■

B.4. **Proof of Corollary 5.** Arbitrarily choose $c \in X$. By A.1, A.2, and Theorem 3 of Luce [11], it follows that

$$p_t(a \mid A) = \frac{r_t(a, c)}{\sum_{b \in A} r_t(b, c)} \delta_a(A) \qquad \forall (t, a, A) \in (0, \infty) \times X \times \mathcal{A}$$

Now set $v_t(a) = \ln r_t(a, c)$ for all $(t, a) \in (0, \infty) \times X$, so that

$$p_t(a \mid A) = \frac{e^{v_t(a)}}{\sum_{b \in A} e^{v_t(b)}} \delta_a(A) \qquad \forall (t, a, A) \in (0, \infty) \times X \times \mathcal{A}$$

Arbitrarily choose $s \in (0, \infty)$, then A.10 implies

$$\begin{aligned} t \ln r_t (a, c) &= s \ln r_s (a, c) \qquad \forall (t, a) \in (0, \infty) \times X \\ v_t (a) &= \frac{s v_s (a)}{t} \qquad \forall (t, a) \in (0, \infty) \times X \end{aligned}$$

Recall s is fixed and define, for all $a \in A$, $E(a)$ by

$$s v_s (a) = -\frac{E(a)}{k}$$

that is,

$$E(a) = -k s v_s (a) = -k s \ln r_s (a, c) = k s \ln r_s (c, a)$$

to obtain

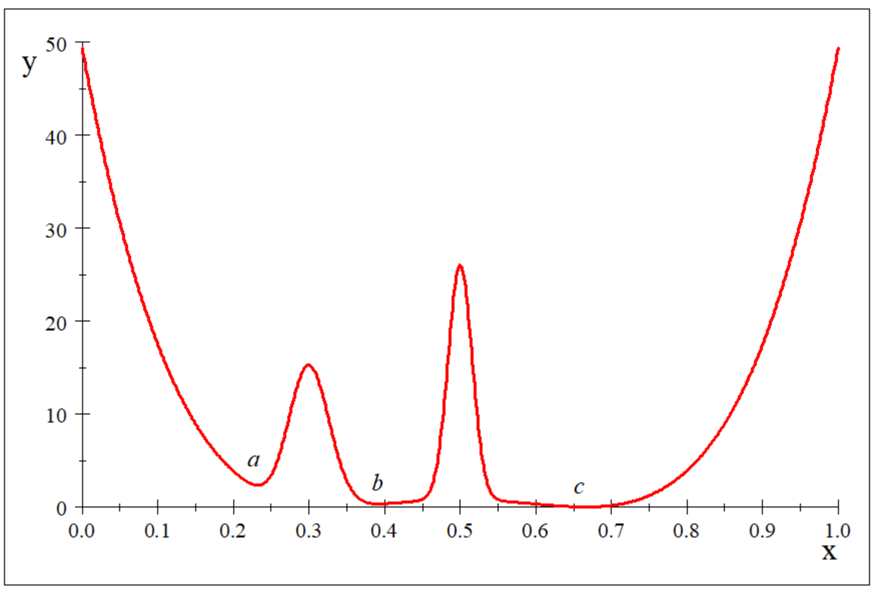
$$v_t (a) = \frac{s v_s (a)}{t} = -\frac{E(a)}{kt} \qquad \forall (t, a) \in (0, \infty) \times X$$

The rest is standard verification. Also note that the same proof continues to hold when $(0, \infty)$ is replaced by any set T of positive temperatures. ■

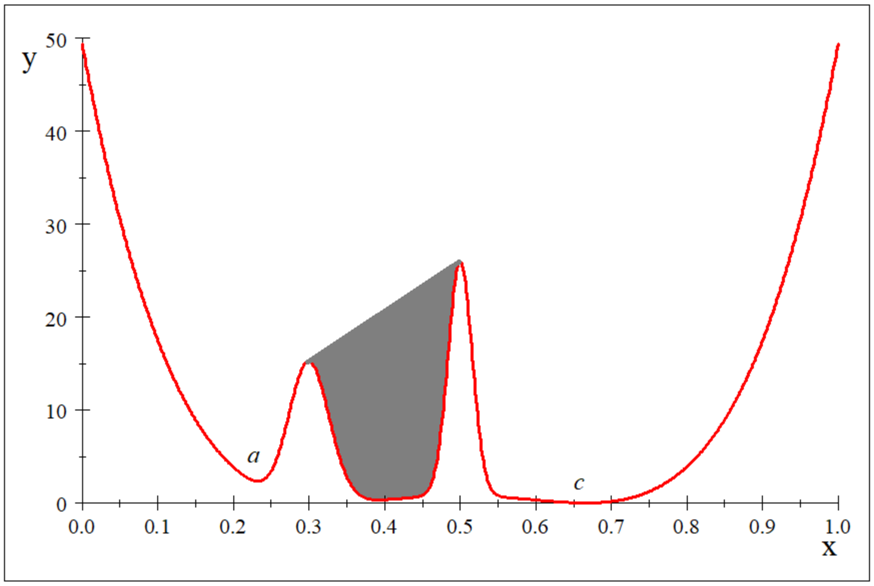
Appendix C. A Violation of Axiom A.2

This appendix presents a counterexample to the Conditioning axiom.

Suppose balls are falling from a random abscissa and ordinate t in the "vase" below. The set of states is the set of local minima $a < b < c$.



When t is not too extreme it is reasonable to expect $p_t (a \mid \{a, b, c\}) / p_t (c \mid \{a, b, c\}) \approx 3/5$ because most balls will stop in the "valley" in which they fall. When state b is excluded by "filling" the vase,



then, it will most likely be the case that $p_t(a \mid \{a,c\})/p_t(c \mid \{a,c\}) \approx 5/5 = 1$ which clearly violates Conditioning.

Università Bocconi
Email address: simone.cerreia@unibocconi.it

Università Bocconi
Email address: fabio.maccheroni@unibocconi.it

Università Bocconi
Email address: massimo.marinacci@unibocconi.it

University of Minnesota
Email address: aldo.rustichini@gmail.com