UNIVERSITÀ DEGLI STUDI DI UDINE

Scuola Superiore di Udine

End Year Thesis

Symmetry as a rational principle

TUTOR:
Prof. Pietro Corvaja

STUDENT:
Michele Mignani

Academic year 2023/2024

Introduction

Modeling knowledge and rationality was one of the reasons why logic was developed starting from Aristotle. Passing through mathematizations of reasoning like Propositional Logic or First Order Logic, there were also attempts to use modal logic to deal with the distance between what is true in the world and what the agent knows about it. For instance, in *Epistemic Logics* some modal operators that represent the knowledge of an agent (or what every agent knows) are added and a Kripke-style semantics is developed (see for further details [18] and [10]).

Even if the modal approach sheds important light on the qualitative study of uncertainty, probability is the tool used from the beginning to investigate the question quantitatively. However, this doesn't mean that what we will present is the perfect approach for dealing with these concepts. Indeed:

- The probabilistic approaches have huge limitations: usually the degree of beliefs of an agent are expressed by real numbers and it's not really clear how to retrieve this value (if it exists) from an agent; moreover, in this framework, we force any two events to be comparable in terms of likelihood to the eyes of an agent. So for a given goal, a different point of view about the topic can be more suitable;
- There are a lot of techniques that use probabilities and what is shown in the thesis is only one among all the possible ways to formalize the question.

Rudolf Carnap's works ([1] among others) are considered one of the main texts about Inductive Logic; with time, the philosophical interests that were at the base of his research began to fade in favor of a more mathematical approach to the topic. Haim Gaifman ([7]), Marcus Hutter ([12]), Jeffrey Paris, and Alena Vencovská ([20], [11], [21], [22], [19]), among others, developed a sound mathematical theory for rational beliefs based on first-order logic and probabilities of sentences. This line of research is called *Inductive Logic* and the main results discovered until now are collected in [20].

The main question that Inductive Logic tries to answer is: "What *degree of belief* should a rational agent have about a certain *property* relative to the *world* it lives in?". Giving a precise meaning to the italic terms is essential to understanding what it really means:

- *degree of belief*: an agent will assign to any *property* a value in $[0, 1]$ depending on how much likely it is: 0 if it is regarded as impossible, 1 if certain;
- *property*: an agent will judge statements expressible in a given logic and a certain language $\mathcal{L}$;[1]

[1] sometimes called *signature*.

- *world*: an agent will be supposed to "live" in a $\mathcal{L}$-structure.

The beliefs of an agent are represented as a map (called *probability*) from the set of all the $\mathcal{L}$-sentences, denoted by $\mathrm{Sen}(\mathcal{L})$, to $[0,1]$. Not all the maps can describe what a *rational* agent thinks and the first part of Chapter 1 is devoted to a more precise discussion about what we mean with *rationality*. There, we will present three different approaches and it will be shown that they all justify the same definition (Definition 1.1.1).

In Chapter 1 we investigate the first symmetry principles ending with the ultimate Invariance Principle: there is only one probability that satisfies this property and sadly this way of assigning beliefs has severe faults that are investigated.

The work will move towards other principles different from symmetry that can be used to study the rationality of a probability: in particular, in Chapter 2 we will investigate the notions of *Relevance* and *Irrelevance*, ending with the presentation of the Carnap's continuum, a family of probabilities depending on a parameter $\lambda \in [0, +\infty]$ that satisfy most of the rational requirements usually accepted by the scholars.

Notations

Throughout the work, we will use a standard notation for usual mathematical objects. However, since we will heavily use logical concepts and they are sometimes represented in different ways in literature, we want to outline clearly the meaning of some symbols.

In the following, we will consider languages $\mathcal{L}$ that are countable, i.e. with a countable supply of variables ($x_1, \dots, x_n, \dots$ or $y_1, \dots, y_n, \dots$ or simply x, y, z), with a countable supply of constants ($\mathrm{Con} = \{a_1, \dots, a_n, \dots\}$) predicate symbols ($R_1, \dots R_q$) that generally are considered unary, if not otherwise specified; we will not consider languages with equality.

The logic that we will use is the first-order classical one (FO) and the set of all sentences (i.e. formulas without free variables) that can be built with the boolean connectives ($\wedge, \vee, \neg, \rightarrow, \equiv$), the first-order quantifiers ($\exists, \forall$), and with the non-logical symbols of $\mathcal{L}$ will be denoted with $\mathrm{Sen}(\mathcal{L})$; the set $\mathrm{QFSen}(\mathcal{L})$ will be the set of all the quantifier-free formulas of $\mathrm{Sen}(\mathcal{L})$. Regarding substitutions, $\varphi\{x/t_i\}$ will also be denoted by $\varphi(t_i)$ when the variable which we are replacing is clear from the context or irrelevant.

We will also use some concepts or results from Model Theory: a generic structure will be denoted by gothic letters such as $\mathfrak{A}, \mathfrak{B}, \dots$. A structure $\mathfrak{A}$ will be called a model of $\varphi \in \mathrm{Sen}(\mathcal{L})$ if φ holds in $\mathfrak{A}$ (denoted by $\mathfrak{A} \vDash \varphi$): often, following most of the literature, the terms "structure" and "model" are used with no difference in meaning, and we will say, for instance, "consider a model $\mathfrak{A}$"even if there is no reference to a sentence. Dealing with formulas φ that have free variables, a structure $\mathfrak{A}$ isn't enough to give meaning to a notion of validity in the model. We need a *state assignment* σ, that is a map from the set of variables in $\mathcal{L}$ to the domain of $\mathfrak{A}$: since the only important values to understand if $\mathfrak{A}, \sigma \vDash \varphi$ are the ones assumed on the free variables occurring in φ, sometimes the domain of σ is considered to be restricted to them. In addition, given $\varphi = \varphi(x_1, \dots, x_n)$, where $x_1, \dots, x_n$ are the variables that appear free in the formula at issue, and given $d_1, \dots, d_n$ elements of the domain of $\mathfrak{A}$, the notation "$\mathfrak{A} \vDash \varphi[d_1, \dots, d_n]$" will mean that given a state assignment σ[2] such that for any $i = 1, \ldots, n$ σ maps x_i to d_i, $\mathfrak{A}, \sigma \models \varphi$.

In general, when new or less common mathematical objects are introduced, we will discuss also the notation used.

[2]By Model Theory results, it is equivalent to say that "for any state assignment σ ..."

Contents

Chapter 1

Symmetry

1.1 First Definitions

In this chapter, after recalling the setting of Pure Inductive Logic we are working with, we analyze the role of symmetry in establishing rational principles.

In our context the rational beliefs of an agent is formalized by the following definition. Recall that, as already mentioned in the Introduction, we will work with languages $\mathcal{L}$ with a countable supply of variables, of constants (denoted by $\mathrm{Con} = \{a_n \ : \ n \in \mathbb{N}^+\}$), neither functional symbols, nor equality.

Definition 1.1.1. Given a language $\mathcal{L}$, a *probability* on $\mathcal{L}$ is a map $w : \mathrm{Sen}(\mathcal{L}) \to \mathbb{R}_{\geq 0}$ such that:

(a) for every valid φ, $w(\varphi) = 1$;

(b) if $\neg(\varphi \wedge \psi)$ is valid, then $w(\varphi \vee \psi) = w(\varphi) + w(\psi)$;

(c) for every φ with one and only one free variable x, we have:

$$w(\exists x \varphi) = \sup\{w(\bigvee_{i \in I} \varphi\{x/a_i\}) \ : \ I \text{ finite subset of } \mathbb{N}\}. \tag{1.1}$$

Example 1.1.1. We can consider a language $\mathcal{L}$ as above specified and a $\mathcal{L}$-structure $\mathfrak{A} \in \mathrm{Mod}_{\mathrm{Con}}(\mathcal{L})$, i.e. an $\mathcal{L}$-structure with the given domain Con with the natural interpretation of the constant symbols, i.e. for any $i \in \mathbb{N}^+$ $a_i^{\mathfrak{A}} = a_i$. We can define the probability $w_{\mathfrak{A}}$ that assigns 1 to all the sentences true in $\mathfrak{A}$ and 0 otherwise. This is a probability since:

- if φ is valid, then it is true in $\mathfrak{A}$ and, therefore, $w_{\mathfrak{A}}(\varphi) = 1$;
- let φ and ψ be in contradiction, i.e. $\neg(\varphi \wedge \psi)$ be valid. Then $\mathfrak{A} \models \neg(\varphi \wedge \psi)$ and thus in $\mathfrak{A}$ only one among φ and ψ can be true. We have two cases:
  - if $\varphi \vee \psi$ is true in $\mathfrak{A}$, then $w_{\mathfrak{A}}(\varphi \vee \psi) = 1$. Furthermore, at least one among φ and ψ is true in $\mathfrak{A}$ and, by the previous remark, exactly one of them is, from which the condition b) of Definition 1.1.1 is satisfied;
  - otherwise, $\varphi \vee \psi$ is not true in $\mathfrak{A}$ and so is φ and ψ. Therefore, $w_{\mathfrak{A}}(\varphi \vee \psi) = w_{\mathfrak{A}}(\varphi) = w_{\mathfrak{A}}(\psi) = 0$, hence the thesis;

– since the domain of $\mathfrak{A}$ is Con, for any φ formula with x as the only free variable, we have that $\mathfrak{A} \vDash \exists x\varphi$ if and only if $\mathfrak{A} \vDash \varphi(a_i)$ for some $i \in \mathbb{N}^+$: from this, the thesis follows.

The just defined $w_{\mathfrak{A}}$ is a truth-values valuation: therefore, this example shows that all these valuations are nothing but probabilities on $\mathrm{Sen}(\mathcal{L})$.

Example 1.1.2. Another example of probability is given by taking convex combinations of the probabilities in Example 1.1.1: given some $\mathcal{L}$-structures $\mathfrak{A}_1, \dots, \mathfrak{A}_n$ and some non-negative reals $m_1, \dots, m_n$ that sum to 1, we can define the probability

$$w = \sum_{i=1}^{n} m_i w_{\mathfrak{A}_i}.$$

We can generalize the example above in the case of a countable quantity of structures: we can choose some structures $\{\mathfrak{A}_i\}_{i\in\mathbb{N}}$, some non-negative weights $\{m_i\}_{i\in\mathbb{N}}$ such that $\sum_{i\in\mathbb{N}} m_i = 1$ and define a new probability $w = \sum_{i\in\mathbb{N}} m_i w_{\mathfrak{A}_i}$. There is no need to take into account only probabilities of the form $w_{\mathfrak{A}_i}$: every convex combination (also countable) of probabilities is a probability.

Perhaps, the reader has already encountered such examples of probabilities in logic expressivity results. For instance, suppose that $\mathcal{L}$ is a finite language without functional symbols. Since[1], for fixed $n \in \mathbb{N}^+$, there are a finite number of structures of cardinality n up to isomorphism, we can define for all $n > 0$ the (finite) set $\mathcal{C}_n$ of all the $\mathcal{L}$-structures with n elements in the domain (up to isomorphisms). For every formula φ, let

$$w_n(\varphi) := \frac{|\{\mathfrak{A} \in \mathcal{C}_n : \mathfrak{A} \vDash \varphi\}|}{|\mathcal{C}_n|}.$$

These w_n's are probabilities in the sense of Definition 1.1.1 and they are linear combinations (with all the weights equal to $\frac{1}{|\mathcal{C}_n|}$) of probabilities of the form $w_{\mathfrak{A}}$ for an $\mathcal{L}$-structure $\mathfrak{A}$.

The limit as n approaches to infinity of these probabilities exists and it is also a probability, a trivial one, though: the $0 - 1$ Law for first-order logic states that for every formula $\varphi \in \mathrm{Sen}(\mathcal{L})$ the limit at issue exists (usually it is called *asymptotic probability*) and the possible outcomes are only 0 or 1.

Example 1.1.3. The many-valued propositional valuations G_k defined by Gödel are not examples of probabilities over sentences for $k > 2$. Even if they are usually defined only for quantifier-free formulas, we present an argument that shows that each of their extension to $\mathrm{Sen}(\mathcal{L})$ can't be a probability, since the violation of Definition 1.1.1 happens already for sentences in $\mathrm{QFSen}(\mathcal{L})$.
We will show it in the case $k = 3$, but it's easy to generalize the result to all the other cases. Let $v : \mathrm{QFSen}(\mathcal{L}) \to \{0, \frac{1}{2}, 1\}$ be the propositional valuation defined inductively on

[1] The absence of functional symbols is necessary for the proof of the theorem we will discuss later in the example; we don't need this assumption to have only finitely many not isomorphic structures with a given cardinality, though.

the height of the formulas following the rules

$$v(\varphi \wedge \psi) = \min\{v(\varphi), v(\psi)\};$$
$$v(\varphi \vee \psi) = \max\{v(\varphi), v(\psi)\};$$
$$v(\neg\varphi) = \begin{cases} 1 & \text{if } v(\varphi) = 0 \\ 0 & \text{otherwise} \end{cases};$$
$$v(\varphi \to \psi) = \begin{cases} 1 & \text{if } v(\varphi) \leq v(\psi) \\ v(\psi) & \text{otherwise} \end{cases}.$$

Considering the case in which v doesn't assume only the values $0, 1$, there is a formula φ for which $v(\varphi) = \frac{1}{2}$. Hence, $v(\neg\varphi) = 0$ and $v(\varphi \vee \neg\varphi) = \frac{1}{2}$, even if $\varphi \vee \neg\varphi$ is a tautology.[2]

Remark 1.1.1. Condition c) in Definition 1.1.1 is called *Gaifman Condition*: it is essentially based on the fact that for the applications we have in mind, the universe is exhausted by the constants of our language, i.e. all the objects of the world the agent is living can be represented by a constant symbol in the language $\mathcal{L}$ used.

The conditions in Definition 1.1.1 are sufficient to guarantee other properties of a probability, as the following proposition shows.

Proposition 1.1.1. *Let w be a probability over a language $\mathcal{L}$. Then for every φ, ψ and $\varphi_i \in \mathrm{Sen}(\mathcal{L})$ the following hold:*

1) $w(\neg\varphi) = 1 - w(\varphi)$;

2) $w(\varphi) \in [0, 1]$;

3) if φ is unsatisfiable, then $w(\varphi) = 0$;

4) if $\varphi \to \psi$ is valid, then $w(\varphi) \leq w(\psi)$;

5) if $\varphi \equiv \psi$ is valid, then $w(\varphi) = w(\psi)$;

6) $w(\varphi \wedge \psi) \leq w(\psi)$ and the equality holds if φ is valid;

7) $w(\bigvee_{i=1}^{n} \varphi_i) \leq \sum_{i=1}^{n} w(\varphi_i)$ and the two terms are equal if the sentences $\varphi_1, \ldots, \varphi_n$ are pairwise contradictory sentences (i.e. for every distinct $i, j \in \{1, \ldots, n\}$, $\varphi_i \wedge \varphi_j$ is unsatisfiable);

8) $w(\varphi \vee \psi) + w(\varphi \wedge \psi) = w(\varphi) + w(\psi)$.

9) if $w(\psi) > 0$, the map $w(-|\psi) : \mathrm{Sen}\,\mathcal{L} \to \mathbb{R}_{\geq 0}$ defined by

$$w(\varphi|\psi) = \frac{w(\varphi \wedge \psi)}{w(\psi)},$$

is a probability;

[2] This result should not surprise the reader, since this kind of semantics was originally developed to study intuitionistic logic: the violation of Definition 1.1.1 we presented, indeed, relies on the fact that the principle of excluded middle is not a tautology according to these valuations.

10) if φ is a formula with x as the only free variable,

$$w(\exists x\varphi) = \lim_{n\to+\infty} w(\bigvee_{i\leq n} \varphi\{x/t_i\}); \tag{1.2}$$

11) if φ is a formula with x as the only free variable,

$$w(\forall x\varphi) = \inf\{w(\bigwedge_{i\in I} \varphi\{x/t_i\}) \ : \ I \text{ finite subset of } \mathbb{N}\} \tag{1.3}$$

12) if φ is a formula with x as the only free variable,

$$w(\forall x\varphi) = \lim_{n\to+\infty} w(\bigwedge_{i\leq n} \varphi\{x/t_i\}) \tag{1.4}$$

Proof. 1) Since $\neg(\varphi \wedge \neg\varphi)$ is valid, then $w(\varphi \vee \neg\varphi) = w(\varphi) + w(\neg\varphi)$. The first term of the precedent equality is 1 because $\varphi \vee \neg\varphi$ is valid, then the thesis;

2) By Definition 1.1.1, $w(\varphi) \geq 0$ but also $w(\neg\varphi) \geq 0$. Thanks to 1), we have $w(\varphi) = 1 - w(\neg\varphi) \leq 1$;

3) It follows from 1) and the fact that the negation of an unsatisfiable formula is a valid one;

4) If $\varphi \to \psi$ is valid, then so $\psi \vee \neg\varphi$ is and therefore $\neg(\neg\psi \wedge \varphi)$ is. By b) of Definition 1.1.1,

$$w(\neg\psi \vee \varphi) = w(\neg\psi) + w(\varphi)$$

and therefore, using 1) and 2)

$$1 \geq w(\neg\psi \vee \varphi) = w(\neg\psi) + w(\varphi) = 1 - w(\psi) + w(\varphi),$$

from which the thesis.

5) It follows from 4)

6) By the excluded middle principle, ψ is equivalent to $(\psi \wedge \varphi) \vee (\psi \wedge \neg\varphi)$. Therefore by 5) and by the fact that $\psi \wedge \varphi$ is in contradiction with $\psi \wedge \neg\varphi$, we have

$$w(\psi) = w((\psi \wedge \varphi) \vee (\psi \wedge \neg\varphi)) = w(\psi \wedge \varphi) + w(\psi \wedge \neg\varphi) \geq w(\psi \wedge \varphi).$$

In the case in which φ is valid, then $\psi \wedge \neg\varphi$ is unsatisfiable, and by 3) the last disequality in the expression above is indeed an equality.

7) Assume first that $\varphi_1, \ldots, \varphi_n$ are pairwise contradictory. Then, also $\varphi_1 \vee \varphi_2 \vee \cdots \vee \varphi_{n-1}$ and φ_n are in contradiction and by b) of Definition 1.1.1,

$$w(\varphi_1 \vee \cdots \vee \varphi_n) = w((\varphi_1 \vee \cdots \vee \varphi_{n-1}) \vee \varphi_n) = w(\varphi_1 \vee \cdots \vee \varphi_{n-1}) + w(\varphi_n).$$

Then, it is easy to see that the thesis follows by induction, in the case of pairwise contradictory sentences.

In the general case, we can notice that

$$\bigvee_{i=1}^{n} \varphi_i \equiv \bigvee_{i=1}^{n-1} \varphi_i \vee (\neg \bigvee_{i=1}^{n-1} \varphi_i \wedge \varphi_n)$$

and that $\bigvee_{i=1}^{n-1} \varphi_i$ and $(\neg \bigvee_{i=1}^{n-1} \varphi_i \wedge \varphi_n)$ are in contradiction. Therefore by 5), b) of Definition 1.1.1 and 6), we have

$$\begin{aligned} w(\bigvee_{i=1}^{n} \varphi_i) = w(\bigvee_{i=1}^{n-1} \varphi_i \vee (\neg \bigvee_{i=1}^{n-1} \varphi_i \wedge \varphi_n)) &= w(\bigvee_{i=1}^{n-1} \varphi_i) + w(\neg \bigvee_{i=1}^{n-1} \varphi_i \wedge \varphi_n)) \\ &\leq w(\bigvee_{i=1}^{n-1} \varphi_i) + w(\varphi_n). \end{aligned}$$

As in the previous case, it's easy to see now that a simple induction yields to the thesis.

8) We can notice that $\varphi \vee \psi$ is equivalent to the formula $\varphi \vee (\neg\varphi \wedge \psi)$, that has the two disjuncts in contradiction. Therefore

$$w(\varphi \vee \psi) = w(\varphi \vee (\neg\varphi \wedge \psi)) = w(\varphi) + w(\neg\varphi \wedge \psi). \tag{1.5}$$

Using that ψ is equivalent to $(\varphi \wedge \psi) \vee (\neg\varphi \wedge \psi)$ and that the disjuncts on this formula are in contradiction, we get

$$w(\psi) = w(\varphi \wedge \psi) + w(\neg\varphi \wedge \psi). \tag{1.6}$$

Combining Equation (1.5) and Equation (1.6) we have the thesis.

9) The condition a) of Definition 1.1.1 holds by 6). To show the validity of the condition b), let φ_1 and φ_2 two sentences that are in contradiction. Logically, $(\varphi_1 \vee \varphi_2) \wedge \psi$ is equivalent to $(\varphi_1 \wedge \psi) \vee (\varphi_2 \wedge \psi)$ and the two disjuncts are in contradiction by the hypothesis we are assuming about φ_1 and φ_2. Therefore,

$$w((\varphi_1 \vee \varphi_2) \wedge \psi) = w(\varphi_1 \wedge \psi) + w(\varphi_2 \wedge \psi),$$

and, then,

$$\begin{aligned} w(\varphi_1 \vee \varphi_2 | \psi) &= \frac{w((\varphi_1 \vee \varphi_2) \wedge \psi)}{w(\psi)} = \frac{w(\varphi_1 \wedge \psi) + w(\varphi_2 \wedge \psi)}{w(\psi)} \\ &= w(\varphi_1|\psi) + w(\varphi_2|\psi). \end{aligned}$$

10) Let s be the right-hand-side of Equation (1.1) and l the limit that appears in item 10) of this proposition: we will prove that $s = l$. Combining 8) and 6) of this proposition, we have for all $\varphi_1, \varphi_2 \in \mathrm{Sen}(\mathcal{L})$

$$w(\varphi_1 \vee \varphi_2) \geq w(\varphi_1),$$

and by induction we have that the succession $\{w(\bigvee_{i\leq n} \varphi\{x/a_i\})\}_{n\in\mathbb{N}}$ is increasing, therefore the limit exists and

$$l := \lim_{n\to+\infty} w(\bigvee_{i\leq n} \varphi\{x/a_i\}) = \sup_{n\in\mathbb{N}} w(\bigvee_{i\leq n} \varphi\{x/a_i\}).$$

Since the sets of the form $\{0, 1, \ldots, n\}$ form a subset of all the finite subsets I of $\mathbb{N}$, $s \geq l$. Now, for an arbitrary ε, from the definition of sup, we have a finite $I \subset \mathbb{N}$ such that $s - w(\bigvee_{i \in I} \varphi\{x/a_i\}) \leq \varepsilon$. If n is the maximum natural number in I, then $w(\bigvee_{i \in I} \varphi\{x/a_i\}) \leq w(\bigvee_{i \leq n} \varphi\{x/a_i\})$ and

$$s - w(\bigvee_{i \leq n} \varphi\{x/a_i\}) \leq \varepsilon.$$

Therefore, as n approaches $+\infty$ we get

$$s - l \leq \epsilon$$

and, since ε is arbitrary, $s \leq l$. Hence, $s = l$ and the thesis follows;

11) using 1) of this proposition, De Morgan's laws and properties of sup,

$$\begin{aligned}
w(\forall x \varphi) &= w(\neg \exists x \neg \varphi) = 1 - w(\exists x \neg \varphi) \\
&= 1 - \sup\{w(\bigvee_{i \in I} \neg \varphi\{x/a_i\}) \ : \ I \text{ finite subset of } \mathbb{N}\} \\
&= 1 - \sup\{w(\neg \bigwedge_{i \in I} \varphi\{x/a_i\}) \ : \ I \text{ finite subset of } \mathbb{N}\} \\
&= 1 - \sup\{1 - w(\bigwedge_{i \in I} \varphi\{x/a_i\}) \ : \ I \text{ finite subset of } \mathbb{N}\} \\
&= 1 - (1 - \inf\{w(\bigwedge_{i \in I} \varphi\{x/a_i\}) \ : \ I \text{ finite subset of } \mathbb{N}\}) \\
&= \inf\{w(\bigwedge_{i \in I} \varphi\{x/a_i\}) \ : \ I \text{ finite subset of } \mathbb{N}\}).
\end{aligned}$$

12) we can proceed in an analogous way as in item 10).

□

Remark 1.1.2. Condition 3) of Proposition 1.1.1 is not invertible: it's possible to find a probability w and a satisfiable sentence φ such that $w(\varphi) = 0$. For instance, take a formula φ that is satisfiable but not valid and a model $\mathfrak{A}$ for $\neg\varphi$. Then, using the notation of Example 1.1.1, $w_{\mathfrak{A}}(\varphi) = 0$. With the same argument, we can show that it is also possible to find a probability w and a sentence φ that is not valid, but $w(\varphi) = 1$.

Remark 1.1.3. Condition 7) of Proposition 1.1.1 is not an equivalence: there can be a probability w on $\mathrm{Sen}(\mathcal{L})$ and two statements φ_1, φ_2 not in contradiction with $w(\varphi_1 \vee \varphi_2) = w(\varphi_1) + w(\varphi_2)$. For instance, take two formulas φ_1 and φ_2 such that $\varphi_1 \wedge \varphi_2$ is satisfiable but not valid and a model $\mathfrak{A}$ such that

$$\mathfrak{A} \not\models \varphi_1 \wedge \varphi_2 \quad \mathfrak{A} \models \varphi_1 \quad \mathfrak{A} \not\models \varphi_2.$$

Then,

$$w_{\mathfrak{A}}(\varphi_1 \vee \varphi_2) = 1 = w_{\mathfrak{A}}(\varphi_1) + w_{\mathfrak{A}}(\varphi_2),$$

even if φ_1 and φ_2 are not in contradiction.

Thanks to 9) of Proposition 1.1.1, we can define, whenever $w(\psi) > 0$, the *conditional probability* $w(-|\psi)$ as pointed out above.

Proposition 1.1.1 explains why the maps in Definition 1.1.1 are called probabilities and not simply "measures": indeed, they can be regarded as maps from $\mathrm{Sen}(\mathcal{L})$ to $[0,1]$ and not to all $\mathbb{R}_{\geq 0}$. Even if some similarities with the usual notion of "probability" can be detected, the reader with a more measure-theoretic approach will however have some trouble with the use of this term for such functions: however it can be shown that there is a tight correspondence between probabilities (in the measure-theoretic sense) on $\mathrm{Mod}_{\mathrm{Con}}(\mathcal{L})$ and probabilities as in Definition 1.1.1 (see [20], Chapter 3).

This result allows us to motivate more the definition of probability as a way to express the beliefs of an agent: indeed, if we think about the elements on $\mathrm{Mod}_{\mathrm{Con}}(\mathcal{L})$ as the worlds where an agent can live in, another way to express its beliefs is to consider how much according to the agent is plausible to be in a model: this is formalized in mathematics by the notion of probability and this approach leads us to an equivalent definition to the one of Definition 1.1.1.

There is also another motivation that we briefly present here and that is hystorically called *Dutch Book Argument.* According to this approach, the beliefs of an agent should match its betting prices. An agent believs in φ at least p if and only if it would accept to play with the dealer the bet game_p for a stake $s > 0$: it gains $s(1-p)$ if φ is true and it loses sp otherwise. For any sentence $\varphi \in \mathrm{Sen}(\mathcal{L})$ we can consider the upper bound of all the p's for which game_p is acceptable to the agent and this defines a function Bel that is a way to express the beliefs of the agent:

$$\begin{aligned}\mathrm{Bel} : \mathrm{Sen}(\mathcal{L}) &\to [0,1] \\ \varphi &\mapsto \sup_{p\in[0,1]} \{\mathrm{game}_p \text{ is acceptable for the agent.}\}\end{aligned}$$

In this game-theoretic setting we have a notion of rationality: an agent can be *Dutch-bookable* if and only if it considers acceptable a set of bets that will lead it to a certain loss; an agent is considered rational if it can't be Dutch-bookable.

Also in this case, the definition of the beliefs of an agent meets the formalization used in Definition 1.1.1: a probability w can't be Dutch-booked and if a function Bel can't be Dutch-booked, then it is a probability (see [20], Chapter 5).

Recall that the language we are using is composed of constant symbols $a_1, a_2, \ldots$ and q relation symbols $R_1, \ldots, R_q$ of ariety $r_1, \ldots, r_q$, respectively. For distinct constants $b_1, \ldots, b_m$ coming from the a_i's, a *state description* (or a *state formula*) for $b_1, \ldots, b_m$ is a sentence of the form

$$\Theta = \bigwedge_{i=1}^{q} \bigwedge_{c_1,\ldots,c_{r_i}} \pm R_i(c_1, \ldots, c_{r_i}),$$

where for any $i = 1, \ldots, q$ the $c_1, \ldots, c_{r_i}$ range over all (not necessarily distinct) choices of $b_1, \ldots, b_m$ and $\pm R_i$ is an abbreviation that stands for either R_i or $\neg R_i$. Since for every $i = 1, \ldots, q$, we take into account all the possible choices of r_i elements among the constants we are interested in, a state description describes all it is possible to know about how $b_1, \ldots, b_m$ relates to each other according to the various R_i's. With this notation, we mean also that for every i, a choice $c_1, c_2, \ldots, c_{r_i}$ appears only one time in the state description: therefore, Θ is satisfiable. It will be useful to give also a meaning when

$m = 0$: in this case, no constant symbols will appear in Θ, and we will interpret it as a tautology (since it is the empty conjunction).

Example 1.1.4. If a language has a unary symbol P and a binary symbol R, then a possible state description for the constants a, b is

$$\Theta(a, b) = P(a) \wedge \neg P(b) \wedge R(a, a) \wedge \neg R(a, b) \wedge R(b, a) \wedge R(b, b).$$

State descriptions will be denoted by upper case Greek letter as Θ, Φ, Ψ: sometimes with a slight abuse of notation, we will call also *state descriptions (for* $x_1, \ldots, x_n$*)* formulas $\Theta(x_1, \ldots, x_n)$ such that for any constants $b_1, \ldots, b_n$, the sentence $\Theta\{x_1/b_1, \ldots, x_n/b_n\}$ is a state description for $b_1, \ldots, b_n$.

State descriptions are useful because every quantifier-free sentence is equivalent to a (finite, possibly empty) disjunction of state descriptions (Disjunctive Normal Form Theorem)[3]; furthermore, two different state descriptions are contradictory, hence if we have a sentence $\varphi(\vec{b}) \in \mathrm{QFSen}(\mathcal{L})$ that is equivalent to $\bigvee_{i \in S} \Theta_i(\vec{b})$ for a given finite set S, we have

$$w(\varphi(\vec{b})) = \sum_{i \in S} w(\Theta_i(\vec{b})).$$

To fully describe a probability w, it is sufficient to determine the values given to state descriptions: having these, indeed, we completely describe w on $\mathrm{QFSen}(\mathcal{L})$ and then there is a unique extension to a probability on $\mathrm{Sen}(\mathcal{L})$; such an extension can be provided by resoning inductively on the number of quantifiers in the prenex form of a formula and by using repeatedly item (c) of Definition 1.1.1.

Another useful property that we show here is that we can compute w on a state description involving some constants $b_1, \ldots, b_m$ by simply knowing the values assumed by w on some state descriptions for $b_1, \ldots, b_r$ with $r \geq m$. Indeed, for any state description $\Theta(b_1, \ldots, b_m)$, we have

$$\Theta(b_1, \ldots, b_m) \equiv \bigvee_{\Phi(b_1,\ldots,b_r) \vDash \Theta(b_1,\ldots,b_m)} \Phi(b_1, \ldots, b_r),$$

i.e. $\Theta(b_1, \ldots, b_m)$ is equivalent to the disjunction of all possible state descriptions Φ for $b_1, \ldots, b_r$ that imply Θ. Hence,

$$w(\Theta(b_1, \ldots, b_m)) = \sum_{\Phi(b_1,\ldots,b_r) \vDash \Theta(b_1,\ldots,b_m)} w(\Phi(b_1, \ldots, b_r)). \tag{1.7}$$

We can notice, also, that the semantic consequence relationship $\vDash$ is easy to determine on state descriptions. Indeed, $\Phi(b_1, \ldots, b_r) \vDash \Theta(b_1, \ldots, b_m)$ if and only if $\Phi(b_1, \ldots, b_r)$ restricted to the constants $b_1, \ldots, b_m$ is equal to $\Theta(b_1, \ldots, b_m)$.

We showed that a probability w is determined uniquely by its value on state descriptions. However, to construct a probability starting from the values assigned to state descriptions, we should pay some attention to avoid an overdetermination or a not well-defined function. However, if we assign a value $w(\Theta)$ for any state description in such a

[3]We use the usual convention that an empty disjunction is $\perp$.

way that the following holds:

$$\begin{aligned}
&w(\Theta(b_1,\dots,b_m)) \geq 0 \text{ for every } b_1,\dots,b_m \text{ constants;}\\
&w(\top) = 1;\\
&\text{Equation (1.7) holds for any } r \geq m \text{ and every constant symbols } b_1,\dots,b_r,
\end{aligned} \tag{1.8}$$

there is a unique extension to a probability w first on QFSen($\mathcal{L}$) and then on Sen($\mathcal{L}$) as previously explained, that coincides with the above values on state descriptions.

If a languge is unary, i.e. all the relational symbols $R_1,\dots,R_q$ are unary, a state description $\Theta(b_1,\dots,b_n)$ for the constants $b_1,\dots,b_n$ (among the set $\{a_i\}_{i\in\mathbb{N}^+}$) has the form

$$\Theta(b_1,\dots,b_n) = \bigwedge_{i=1}^{n}\bigwedge_{j=1}^{q} \pm R_j(b_i),$$

where $+R$ and $-R$ are intended to be, respectively, R and $\neg R$. Another representation of a state description is given taking into account *atoms* i.e. formulas $\alpha(x)$ of the form $\bigwedge_{j=1}^{q} \pm R_j(x)$. In the case of unary languages, an atom where x is replaced by a constant b describes all we have to know about b[4]; furthermore, from the assumption that the set of relation symbols is finite (it has cardinality q), it follows that also the number of all the possible atoms is finite (it has cardinality 2^q and they will be denoted $\alpha_1(x),\dots,\alpha_{2^q}(x)$). Any state description, henceforth, can be seen as a conjunction of atoms[5], i.e.

$$\Theta(b_1,\dots,b_n) = \bigwedge_{i=1}^{n} \alpha_{h_i}(b_i),$$

where the h_i's vary in the set $\{1,\dots,2^q\}$.

1.2 Symmetry

In this section, we will focus on the unary case: in general, as always, if not otherwise specified, the language will have constants $\{a_n\}_{n\in\mathbb{N}^+}$, unary relational symbols $R_1,\dots,R_q$, neither functional symbols nor equality.

To motivate the principles we will present, suppose to have an urn with some balls (say for simplicity a countable supply), each with a non-zero natural number on it. Suppose that different balls have different numbers written on them (so, there is one and only one ball for each positive natural) and that these balls can be red or have a different color; in addition, the balls can have a white circle on them or not.

To formalize this situation, we will use a language $\mathcal{L}$ composed of a countable supply of constants $\{a_n\}_{n\in\mathbb{N}^+}$ and two unary predicate symbols $P_1(x)$ and $P_2(x)$.

In the following, we will assume that the agent is in the *zero-knowledge* condition, i.e. it doesn't know any relevant information about the world it lives in. In this setting, this means that the agent knows only that the universe is exhausted by the constant symbols

[4]Notice that this is not the case when the language has at least a binary relation symbol.

[5]Here and in the following we will use the world *atom* not only to mean a formula like $\alpha_i(x)$ with x as free variable but also the statements of the form $\alpha_i(b)$ for a constant symbol b: this is in analogy with the terminology used for state descriptions.

in $\mathcal{L}$ and that the entities it is studying can satisfy two unary properties, represented by the predicate symbols $P_1(x)$ and $P_2(x)$.

We will present some valid formalizations that can be used:

- *Formalization 1:*
  The constant a_i is assigned to the ball in which the number i is written; $P_1(a_i)$ stands for "the ball associated with a_i is red"; $P_2(a_i)$ stands for "the ball associated to a_i has a white circle".

- *Formalization 2:*
  The constant a_i is assigned to the ball in which the number i is written, for every $i \notin \{1, 2\}$; a_1 is assigned to the ball in which it is written 2 and a_2 to the ball in which it is written 1; $P_1(a_i)$ stands for "the ball associated to a_i is red"; $P_2(a_i)$ stands for "the ball associated to a_i has a white circle".

- *Formalization 3:*
  The constant a_i is assigned to the ball in which the number i is written, for every i; $P_1(a_i)$ stands for "the ball associated to a_i isn't red"; $P_2(a_i)$ stands for "the ball associated to a_i has a white circle".

- *Formalization 4:*
  The constant a_i is assigned to the ball in which the number i is written, for every i; $P_1(a_i)$ stands for "the ball associated to a_i has a white circle"; $P_2(a_i)$ stands for "the ball associated to a_i is red".

Formalization 1 and Formalization 4 can seem neater than the others: Formalization 2 involves an assignment of constants to balls that is not the first that comes to mind; in Formalization 3, $P_1(a_i)$ means that the ball a_i is *not* red, even if it seems more natural that $P_1(x)$ stands for "x is red".

However, these formalizations are all valid and expressive in the same way: they differ only in the name given to objects or properties we want to study.

Suppose we want to formalize the following property: the ball with 1 written on it is red. According to the different formalizations, we have different sentences that represent it:

- *Formalization 1:* $P_1(a_1)$;
- *Formalization 2:* $P_1(a_2)$;
- *Formalization 3:* $\neg P_1(a_1)$;
- *Formalization 4:* $P_2(a_1)$.

In general, if a property is represented in Formalization 1 by a formula φ_1, then:

- in Formalization 2 the property is expressed by φ_2 that is the outcome of replacing the occurrences (if any) of a_1 with a_2 and of a_2 with a_1;

- in Formalization 3 the property is expressed by φ_3 that is the outcome of replacing the occurrences (if any) of P_1 with $\neg P_1$;

- in Formalization 4 the property is expressed by φ_4 that is the outcome of replacing the occurrences (if any) of P_1 with P_2 and of P_2 with P_1.

A rational agent should observe this ambiguity of the language: it should not give different values to sentences that can express the same property. In this way, we can justify some principles that appears in literature:

- comparing Formalization 1 and Formalization 2, an agent should consider φ_1 as likely as φ_2. Generalizing this example, we can deal with permutations of $\mathbb{N}^+$ more complex than the one that swaps a_1 and a_2 and get the *Constant Exchangeability Principle (Ex)*: a probability w satisfies Ex if for any statement $\varphi(b_1,\dots,b_n)$ with $b_1,\dots,b_n$ distinct constant symbols, and for any other n distinct constant symbols $b'_1,\dots,b'_n$, we have that $w(\varphi(b_1,\dots,b_n)) = w(\varphi(b'_1,\dots,b'_n))$;
- comparing Formalization 1 and Formalization 3, we get the *Strong Negation Principle (SN)*: a probability w satisfies SN if for any statement φ and for any predicate symbol R, if φ' is the outcome of replacing in φ any occurrence of R with $\neg R$, we have that $w(\varphi) = w(\varphi')$;
- comparing Formalization 1 and Formalization 4, we get the*Predicate Exchangeability Principle (Px) - Unary case*: if R and R' are two relational symbols and for any statement φ, φ' is the outcome of replacing in φ any occurence of R with R' and any occurence of R' with R, then $w(\varphi) = w(\varphi')$.

As it will be more clear in the future, these principles are regarded as *symmetry* ones: Ex reflects a symmetry between constants, Px between predicate symbols (of the same ariety), and SN between a predicate symbol and its negation.

Ex is a principle that is agreed by all the scholars and sometimes the term *probability* denotes a probability as in Definition 1.1.1 satisfying Ex.

The importance of Ex is reflected by the relevance of an essential representation result: the De Finetti's Theorem. In the following the set $\mathbb{D}_{2^q}$ will be the simplex in $\mathbb{R}^{2^q}$

$$\mathbb{D}_{2^q} = \{(x_1,\dots,x_{2^q}) \in [0,1]^{2^q} \ : \ \sum_{i=1}^{2^q} x_i = 1\}.$$

Given $\vec{x} \in \mathbb{D}_{2^q}$, we have a probability on $\mathrm{Sen}(\mathcal{L})$ putting

$$w_{\vec{x}}(\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)) = \prod_{i=1}^{m} x_{h_i} = \prod_{j=1}^{2^q} x_j^{m_j},$$

for all the state descriptions and then extending to all the sentences in $\mathrm{QFSen}(\mathcal{L})$ and in $\mathrm{Sen}(\mathcal{L})$ as provided by previous arguments, where $m_j = |\{i \ : \ h_i = j\}|$. To verify that $w_{\vec{x}}$ defines a probability, we can check whether the conditions required in (1.8) are satisfied:

- for every state descriptions $\Theta = \bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)$,

$$w_{\vec{x}}(\Theta) = \prod_{i=1}^{m} x_{h_i} \geq 0,$$

  since every $\vec{x} \in \mathbb{D}_{2^q}$ has all the components non-negative;

- $w_{\vec{x}}(\top) = 1$ because it is the empty product.[6]
- suppose $r \geq m$ and that we have a state description $\Theta = \bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)$. Then, a state description $\bigwedge_{j=1}^{r} \alpha_{f_j}(b_j)$ has Θ as a semantic consequence if and only if for every $j \leq m$, $h_j = f_j$. This means that

$$\begin{aligned}
&\sum_{\Phi(b_1,\dots,b_r)\models\Theta(b_1,\dots,b_m)} w_{\vec{x}}(\Phi(b_1,\dots,b_r)) = \\
&= \sum_{g:\{m+1,\dots,r\}\to\{1,\dots,2^q\}} w_{\vec{x}}\big(\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i) \wedge \bigwedge_{i=m+1}^{r} \alpha_{g(i)}(b_i)\big) \\
&= \sum_{g:\{m+1,\dots,r\}\to\{1,\dots,2^q\}} \prod_{i=1}^{m} x_{h_i} \cdot \prod_{i=m+1}^{r} x_{g(i)} \\
&= \prod_{i=1}^{m} x_{h_i} \cdot \big(\sum_{g:\{m+1,\dots,r\}\to\{1,\dots,2^q\}} \prod_{i=m+1}^{r} x_{g(i)}\big) \\
&= \prod_{i=1}^{m} x_{h_i} = w_{\vec{x}}\big(\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)\big) = w_{\vec{x}}(\Theta(b_1,\dots,b_m)).
\end{aligned}$$

Then, we can extend this to a probability to all $\mathrm{Sen}(\mathcal{L})$ by using item (c) of Definition 1.1.1, as at page 17.

Theorem 1.2.1 (De Finetti).
Let $\mathcal{L}$ be a unary language with q relation symbols and w a probability on $\mathrm{Sen}(\mathcal{L})$ satisfying Ex. Then there is a unique probability μ on the borel σ-algebra of $\mathbb{D}_{2^q}$ such that

$$w(\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)) = \int_{\mathbb{D}_{2^q}} w_{\vec{x}}(\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i))\ d\mu(\vec{x}),^{7} \tag{1.9}$$

for any $m \in \mathbb{N}^+$ and atoms α_{h_i} of the language.

Conversely, if μ is a probability on the borel σ-algebra of $\mathbb{D}_{2^q}$, and w is defined as in Equation (1.9), then it can be extended to a probability on $\mathrm{Sen}(\mathcal{L})$ that satisfies Ex.

As an immediate corollary, we can generalize Equation (1.9) from conjunctions of atoms to sentences.

Corollary 1.2.2. *With the notations used in Theorem 1.2.1, if w is a probability that satisfies Ex, then there exists a unique probability μ on the borel σ-algebra of $\mathbb{D}_{2^q}$ such that for every $\varphi(b_1,\dots,b_n) \in \mathrm{Sen}(\mathcal{L})$*

$$w(\varphi(b_1,\dots,b_n)) = \int_{\mathbb{D}_{2^q}} w_{\vec{x}}(\varphi(b_1,\dots,b_n))\ d\mu(\vec{x}). \tag{1.10}$$

Conversely, if μ is a probability on the borel σ-algebra of $\mathbb{D}_{2^q}$, then, using Equation (1.10), we can define a probability w that satisfies Ex.

---

[6] This is a usual convention. However, if we wanted to avoid this fuzzy notation, we could define directly $w_{\vec{x}}(\top) = 1$.

[7] We can notice that the function $\vec{x} \mapsto w_{\vec{x}}(\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i))$ is measurable since it is a polynomial, hence continuous.

Theorem 1.2.1, clarified by Corollary 1.2.2, provides a useful tool for dealing with probabilities that satisfy Ex. Indeed, such a probability can be seen as a convex combination (according to μ) of probabilities with a specific form, i.e. of $w_{\vec{x}}$'s with $\vec{x} \in \mathbb{D}_{2^q}$; furthermore, the probability μ that determines the "weights of the various points in the simplex"[8] with regards to the representation of w is unique. In the following, given a probability w, we will call such a μ the *De Finetti prior* of w. Let's show an example.

Example 1.2.1. Assume to have a language $\mathcal{L}$ with only the unary relation symbol R and let's describe the probability w starting from its De Finetti prior. If λ is the Lebesgue measure on $[0, 1]$, then the map $f : t \mapsto (t, 1 - t)$ from $[0, 1]$ to $\mathbb{D}_2$ determines the pushforward probability $f_*\lambda$ on $\mathbb{D}_2$. In this case, for any function $g : \mathbb{D}_2 \to \mathbb{R}$, we have

$$\int_{\mathbb{D}_2} g(\vec{x})\ df_*\lambda(\vec{x}) = \int_{[0,1]} g(f(t))\ d\lambda(t).$$

In this language, we have only two atoms $\alpha_1(x) = R(x)$ and $\alpha_2(x) = \neg R(x)$ and for any constant b, w gives the same value to all the atoms:

$$w(R(b)) = \int_{\mathbb{D}_2} x_1\ d\lambda((x_1, x_2)) = \int_{[0,1]} t\ dt = \frac{1}{2}$$
$$w(\neg R(b)) = \int_{\mathbb{D}_2} x_2\ d\lambda((x_1, x_2)) = \int_{[0,1]} 1 - t\ dt = \frac{1}{2}.$$

If we have a state description $\bigwedge_{i=1}^{n+k} \alpha_{h_i}(b_i)$ with $n = |\{i\ :\ h_i = 1\}|$ and $k = |\{i\ :\ h_i = 2\}$,

$$\begin{aligned} w(\bigwedge_{i=1}^{n+k} \alpha_{h_i}(b_i)) &= \int_{\mathbb{D}_2} x_1^n x_2^k\ d\lambda((x_1, x_2)) \\ &= \int_{[0,1]} t^n(1-t)^k\ dt = \frac{k!n!}{(n+k+1)!}. \end{aligned} \tag{1.11}$$

So, since the outcome is the same if we swap n and k, w satisfies SN: this probability, indeed, in some sense, keeps the major symmetry of the Lebesgue measure. We can notice also that when quantified formulas are involved, we get

$$\begin{aligned} w(\forall x R(x)) &= \int_{\mathbb{D}_2} w_{\vec{x}}(\forall x R(x))\ d\lambda((x_1, x_2)) = \int_{\mathbb{D}_2} \lim_{n\to+\infty} w_{\vec{x}}(\bigwedge_{i=1}^{n} R(a_i))\ d\lambda((x_1, x_2)) \\ &= \lim_{n\to+\infty} \int_{\mathbb{D}_2} x_1^n d\lambda((x_1, x_2)) = \lim_{n\to+\infty} \int_0^1 t^n\ dt = \lim_{n\to+\infty} \frac{1}{n+1} = 0. \end{aligned}$$

We end this example by showing a very useful feature of this probability related to epistemic induction. By previous computation, using the same notation used in (1.11), if $n = |\{i\ :\ h_i = 1\}|$ and $k = |\{i\ :\ h_i = 2\}$, we have that

$$\begin{aligned} w(\alpha_1(a_{n+k+1}) | \bigwedge_{i=1}^{n+k} \alpha_{h_i}(a_i)) &= \frac{w(\bigwedge_{i=1}^{n+k} \alpha_{h_i}(a_i) \wedge \alpha_1(a_{n+k+1}))}{w(\bigwedge_{i=1}^{n+k} \alpha_{h_i}(a_i))} \\ &= \frac{(n+1)!k!}{(n+k+2)!} \frac{(n+k+1)!}{n!k!} = \frac{n+1}{n+k+2}. \end{aligned}$$

[8] This makes sense if μ is a discrete probability; there can be cases in which this is not, but however this can be thought as an informal description of the role of μ in the representation of w.

This means that an agent with this probability, after having seen a lot of constants for which R holds, is more inclined to believe that also it holds for a new constant a_{n+k+1}: for instance, roughly speaking, after having seen 10 objects, 9 of which satisfies R, the probability it assigns to the chance that a new object satisfies $R(x)$ is 10/12.[9]

There is an analogous result also for polyadic languages (see [26] for a detailed derivation) that we omit here.

Following the notion of symmetry suggested by Ex, SN, and Px, as we will see, we can justify other principles; here, we will present maybe the first that is possible to encounter besides the ones previously listed: the *Atom Exchangeability Principle (Ax)*. A probability w satisfies Ax if for any function $h : \{1, \dots, n\} \to \{1, \dots, 2^q\}$, for any permutation τ of $\{1, \dots, 2^q\}$ and for any constants $b_1, \dots, b_n$,

$$w\big(\bigwedge_{i=1}^{n} \alpha_{h_i}(b_i)\big) = w\big(\bigwedge_{i=1}^{n} \alpha_{\tau(h_i)}(b_i)\big).$$

where, as usual, $\alpha_{h_i}(x), \alpha_{\tau(h_i)}(x)$ are atoms.

If we want to exploit symmetry to give rise to many other rational principles, we should define what we mean by it. In other branches of mathematics, the notion of symmetry is extremely linked to the one of automorphism: the more automorphisms an object (a geometrical figure, an algebraic group) has, the more is regarded as symmetric. Dealing with Inductive Logic, it's not very clear what should be the "object" to consider.

Recall that an agent can have symmetry only in a zero-knowledge context: if it already knows that $P(a_1) \wedge \neg P(a_2)$ holds, we can't ask it to deal with a_1 and a_2 in the same way. In this condition, an agent knows only a few things: the logical rules, that the universe entities are exhausted by the set Con, that the properties of these are expressed by means of the relational symbols in $\mathcal{L}$, and that it lives in one structure in $\mathrm{Mod}_{\mathrm{Con}}(\mathcal{L})$. Hence, in the literature, it is proposed the following definition of automorphism.

Definition 1.2.1. Consider a language $\mathcal{L}$ with $\{a_n\}_{n\in\mathbb{N}^+}$ as the set of constants and the two-sorted set $M(\mathcal{L}) := (\mathrm{Mod}_{\mathrm{Con}}(\mathcal{L}), B(\mathcal{L}))$ where,

$$B(\mathcal{L}) = \{\mathrm{Mod}_{\mathrm{Con}}(\varphi) \ : \ \varphi \in \mathrm{Sen}(\mathcal{L})\}.$$

An *automorphism* of $M(\mathcal{L})$ is a bijective map $\gamma : \mathrm{Mod}_{\mathrm{Con}}(\mathcal{L}) \to \mathrm{Mod}_{\mathrm{Con}}(\mathcal{L})$ such that for each $\varphi \in \mathrm{Sen}(\mathcal{L})$, there is a $\psi \in \mathrm{Sen}(\mathcal{L})$ (that will be named $\gamma(\varphi)$) for which

$$\gamma(\mathrm{Mod}_{\mathrm{Con}}(\varphi)) := \{\gamma(\mathfrak{A}) \ : \ \mathfrak{A} \in \mathrm{Mod}_{\mathrm{Con}}(\varphi)\} = \mathrm{Mod}_{\mathrm{Con}}(\psi) = \mathrm{Mod}_{\mathrm{Con}}(\gamma(\varphi)). \quad (1.12)$$

Remark 1.2.1. The requirement expressed in (1.12) is equivalent to the following: for any $\varphi \in \mathrm{Sen}(\mathcal{L})$ and for any $\mathfrak{A} \in \mathrm{Mod}_{\mathrm{Con}}(\mathcal{L})$,

$$\mathfrak{A} \models \varphi \text{ if and only if } \gamma(\mathfrak{A}) \models \gamma(\varphi). \quad (1.13)$$

Indeed, assuming (1.12), we get (1.13), since:

- given a model $\mathfrak{A}$ of φ, then since $\gamma(\mathrm{Mod}_{\mathrm{Con}}(\varphi)) = \mathrm{Mod}_{\mathrm{Con}}(\gamma(\varphi))$, $\gamma(\mathfrak{A})$ satisfies $\gamma(\varphi)$;

[9] This probability is usually called $c^{\mathcal{L}}_{2^q}$: we are using here a standard notation attributed to Carnap and that will be clear after the Definition 2.2.1 in Chapter 2.

- if for a structure $\mathfrak{A}$, $\gamma(\mathfrak{A}) \vDash \gamma(\varphi)$, then there is a model $\mathfrak{B}$ of φ such that $\gamma(\mathfrak{A}) = \gamma(\mathfrak{B})$. From the injectivity of γ, we get that $\mathfrak{A} = \mathfrak{B}$ is a model of φ itself.

Conversely, given (1.13), we can prove (1.12):

- $\gamma(\mathrm{Mod}_{\mathrm{Con}}(\varphi)) \subseteq \mathrm{Mod}_{\mathrm{Con}}(\gamma(\varphi))$: given a model $\mathfrak{A}$ of φ, by (1.13), $\gamma(\mathfrak{A}) \vDash \gamma(\varphi)$;
- $\gamma(\mathrm{Mod}_{\mathrm{Con}}(\varphi)) \supseteq \mathrm{Mod}_{\mathrm{Con}}(\gamma(\varphi))$: given $\mathfrak{B} \in \mathrm{Mod}_{\mathrm{Con}}(\gamma(\varphi))$, for the surjectivity of γ, there exists $\mathfrak{A}$ such that $\gamma(\mathfrak{A}) = \mathfrak{B}$. By (1.13), since $\mathfrak{B} \vDash \gamma(\varphi)$, $\mathfrak{A}$ satisfies φ.

We will present some class of automorphisms of $M(\mathcal{L})$ to get the reader acquainted with this notion.

- *Automorphism that permutes contants:*
  Given a permutation σ of $\mathbb{N}^+$ we can construct the following map $\gamma : \mathrm{Mod}_{\mathrm{Con}}(\mathcal{L}) \to \mathrm{Mod}_{\mathrm{Con}}(\mathcal{L})$: for any structure $\mathfrak{A}$, $\gamma(\mathfrak{A})$ is the structure with Con as domain and the following interpretation of any predicate symbol R
  $$\gamma(\mathfrak{A}) \vDash R(a_i) \text{ if and only if } \mathfrak{A} \vDash R(a_{\sigma(i)}),$$
  for any $i \in \mathbb{N}^+$. This map is bijective, since σ is invertible and we can perform the same construction with σ^{-1}; furthermore, for any $\varphi(a_{i_1}, \ldots, a_{i_n}) \in \mathrm{Sen}(\mathcal{L})$, we have that
  $$\gamma(\mathfrak{A}) \vDash \varphi(a_{i_1}, \ldots, a_{i_n}) \text{ if and only if } \mathfrak{A} \vDash \varphi(a_{\sigma(i_1)}, \ldots, a_{\sigma(i_n)}),$$
  hence
  $$\gamma(\mathrm{Mod}_{\mathrm{Con}}(\varphi(a_{\sigma(i_1)}, \ldots, a_{\sigma(i_n)}))) = \mathrm{Mod}_{\mathrm{Con}}(\varphi(a_{i_1}, \ldots, a_{i_n}))$$
  and putting $\gamma(\varphi(a_{i_1}, \ldots, a_{i_n})) = \varphi(a_{\sigma^{-1}(i_1)}, \ldots, a_{\sigma^{-1}(i_n)})$, we get an automorphism of $M(\mathcal{L})$.

- *Automorphism that permutes predicate symbols:*
  In the unary case, the predicate symbols $R_1, \ldots, R_q$ have the same ariety. For any σ permutation of $\{1, \ldots, q\}$, we can construct the following map $\gamma : \mathrm{Mod}_{\mathrm{Con}}(\mathcal{L}) \to \mathrm{Mod}_{\mathrm{Con}}(\mathcal{L})$: for any structure $\mathfrak{A}$, $\gamma(\mathfrak{A})$ is the structure with Con as domain and the following interpretation of predicate symbols
  $$\gamma(\mathfrak{A}) \vDash R_i(b) \text{ if and only if } \mathfrak{A} \vDash R_{\sigma(i)}(b),$$
  for any $b \in \mathrm{Con}$. Then, arguing in a similar way as the previous point, we can show that γ can be extended to an automorphism of $M(\mathcal{L})$;

- *Automorphism that negates the occurrences of a predicate symbol:*
  For any $i \leq q$, we can construct the following map $\gamma : \mathrm{Mod}_{\mathrm{Con}}(\mathcal{L}) \to \mathrm{Mod}_{\mathrm{Con}}(\mathcal{L})$: for any structure $\mathfrak{A}$, $\gamma(\mathfrak{A})$ is the structure with Con as domain such that the interpretations of R_j with $j \neq i$ is the same as in $\mathfrak{A}$, but
  $$\gamma(\mathfrak{A}) \vDash R_i(b_1, \ldots, b_n) \text{ if and only if } \mathfrak{A} \vDash \neg R_i(b_1, \ldots, b_n),$$
  for any $b_1, \ldots, b_n \in \mathrm{Con}$. Notice that in this way we have also that
  $$\gamma(\mathfrak{A}) \vDash \neg R_i(b_1, \ldots, b_n) \text{ if and only if } \mathfrak{A} \vDash R_i(b_1, \ldots, b_n).$$
  As before, γ extends to an automorphism of $M(\mathcal{L})$.

The invariance of a probability w under these classes of automorphisms is a valuable property: indeed, it corresponds to requiring w to satisfy, respectively, Ex, Px (in the unary case), and SN. We will state this more precisely after the definition of the *invariance* principle.

Definition 1.2.2. A probability w satisfies the *invariance principle* (INV) if and only if for any γ automorphism of the two-sorted structure $M(\mathcal{L})$ and for any $\varphi \in \mathrm{Sen}(\mathcal{L})$, $w(\varphi) = w(\gamma(\varphi))$.

Proposition 1.2.1. *In the unary case, INV implies Ex, Px, SN.*

Proof. We have already proved this by exhibiting the classes of automorphisms the inviariance under which is equivalent to the wanted principles. □

Proposition 1.2.1 shows us that all the formalization-based argument presented at the beginning of the chapter is encompassed by INV. However, what was said is not able to justify the whole Invariance Principle, that is, in our opinion, not reasonable. Our thought about INV doesn't rely only on the absence of motivations providing grounds for it, but also on what is implied by this principle.

In the unary case, INV generates also other symmetry principles, not only the ones previously encountered, but also, for instance Ax. Actually it leaves us with a single probability $c_0^{\mathcal{L}}$[10] defined as follows:

$$c_0^{\mathcal{L}}(\bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)) = \begin{cases} 1 & \text{if } n = 0; \\ 2^{-q} & \text{if } n \geq 1 \text{ and } h_1 = h_2 = \cdots = h_n, \\ 0 & \text{otherwise.} \end{cases}$$

To show that this is a probability, we verify the conditions at page 17:

- for any state description $\Theta(b_1, \ldots, b_m)$ we have that the value assigned to it by $c_0^{\mathcal{L}}$ is $0, 2^{-q}$ or 1, hence in any case it is non-negative;
- since we use the convention that an empty conjunction is equivalent to $\top$, $c_0^{\mathcal{L}}(\top) = 1$ by how the map is defined;
- if $m = 0$, then $\Theta = \top$ and given $r \geq m$, any state description Φ for $b_1, \ldots, b_r$ implies Θ: since there are 2^q state descriptions for which the value given to them by $c_0^{\mathcal{L}}$ is 2^{-q} (one for each atom of the language), and to all the others it is assigned 0, we get that
$$c_0^{\mathcal{L}}(\top) = \sum_{\Phi(b_1,\ldots,b_r) \models \top} c_0^{\mathcal{L}}(\Phi(b_1, \ldots, b_r)) = 2^2 \frac{1}{2^q};$$
- if $m > 0$, taking $r \geq m$, constants $b_1, \ldots, b_r$, and a state description $\Theta(b_1, \ldots, b_m) = \bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)$ we have two cases:
  - if all the h_i's are equal to some $j \in \{1, \ldots, 2^q\}$, $c_0^{\mathcal{L}}(\Theta) = 2^{-q}$. Among all the state descriptions $\Phi(b_1, \ldots, b_r)$ that imply Θ, there is only one ($\Phi = \bigwedge_{i=1}^{r} \alpha_j(b_i)$) for which the value assigned by $c_0^{\mathcal{L}}$ is different from 0 and it is 2^{-q}: this yields to the thesis;

[10] See Footnote 9, page 22.

– otherwise, suppose that for two indexes $i_0, i_1 \in \{1, \ldots, m\}$ we have that $h_{i_0} \neq h_{i_1}$. In any state description $\Phi(b_1, \ldots, b_r)$ that implies Θ, will appear the conjunct $\alpha_{h_{i_0}}(b_{i_0}) \wedge \alpha_{h_{i_1}}(b_{i_1})$ and Equation (1.7) holds since both the terms of the equality are null.

Theorem 1.2.3. *Given a unary language $\mathcal{L}$, the only probability that satisfies INV is $c_0^{\mathcal{L}}$.*

Proof. For a detailed proof see [26], Chapter 3. □

Even if the previous section was devoted to the unary case, some of the definitions given can be carried out also for a polyadic language: in particular the concept of automoprhism and the principle INV can be easily generalized.

The question of whether INV restricts the range of available probabilities to only one also in the polyadic case, was open for a long time. Recently, however, it was proved that it is so.[11]

The probability we will get at the end is a generalization of $c_0^{\mathcal{L}}$ to the polyadic setting. The definition we gave of $c_0^{\mathcal{L}}$ is essentially based on the concept of atoms, a notion meaningful only in the unary context: however, the properties defining $c_0^{\mathcal{L}}$ can be rephrased in a more generalizable way.

In the unary case, for any atom $\alpha_i(x)$ we get a structure $\mathfrak{A}_i$ in which $\forall x \alpha_i(x)$ holds. We can see $c_0^{\mathcal{L}}$ as the probability on Sen($\mathcal{L}$) such that

$$w = \sum_{i=1}^{2^q} \frac{1}{2^q} w_{\mathfrak{A}_i},$$

with the notation used in Example 1.1.1. The structures $\mathfrak{A}_i$ are the only ones in which all the elements of the domain are indistinguishable: the whole behavior of a constant $b \in$ Con is represented by the atom it satisfies when the language is unary. Hence, if $x \sim y$ is the formula

$$\bigwedge_{i=1}^{2^q} (\alpha_i(x) \equiv \alpha_i(y)),$$

the $\mathfrak{A}_i$'s are the different models of $\forall x \forall y\ x \sim y$.

Considering a polyadic language $\mathcal{L}$ that, for simplicity of notation, we will suppose to have only a binary predicate R, $x \sim y$ is the formula

$$\forall z (R(x, z) \equiv R(y, z) \wedge R(z, x) \equiv R(z, y)).$$

Then we have the structures $\mathfrak{B}_0$ and $\mathfrak{B}_1$ such that

$$\mathfrak{B}_0 \models \forall x \forall y (x \sim y \wedge R(x, x)) \quad \mathfrak{B}_1 \models \forall x \forall y (x \sim y \wedge \neg R(x, x)),$$

and the probability ω that generalizes $c_0^{\mathcal{L}}$ to a polyadic language $\mathcal{L} = \{R\}$ is

$$\omega = \frac{1}{2} w_{\mathfrak{B}_0} + \frac{1}{2} w_{\mathfrak{B}_1}.$$

Theorem 1.2.4. *ω is the only probability that satisfies INV.*

[11]This information is based on personal communication with J. Paris and A. Vencovská: the results are under publications ([19]).

1.3 What next?

In the previous section, we saw that the INV principle leaves us with a single probability. The golden goal of this branch of logic is finding rational principles to establish how much an agent should believe in a fact, i.e. in a logical formula that formalizes the fact: thus, have we pursued our aim?

The fact that we have found only one available probability consistent with INV seems to positively answer the question. Actually, this is not the case at all. Dealing with the unary case, indeed, $c_0^{\mathcal{L}}$ has some unpleasant faults; we will see this using the example of the urn full of balls used along the work:

- $c_0^{\mathcal{L}}$ generalizes too much: if the agent discovers that a ball is red, then all the balls are red according to it. This relies on the fact that for any relational symbol P of our language, $c_0^{\mathcal{L}}(-|P(a_i))$ is a probability that gives 0 to all the sentences of the form $\neg P(a_j)$, for all $i, j \in \mathbb{N}^+$;
- $c_0^{\mathcal{L}}$ can't be used in the process of learning: suppose that the agent discovers that a ball is red and another one is not. Then, assuming we are using the natural Formalization 1 of the problem, the agent finds out that there are $i, j \in \mathbb{N}^+$ such that the sentence $P_1(a_i) \wedge \neg P_1(a_j)$ is true. If we want to adjust the prior probability after this observation, we should take the conditional probability $c_0^{\mathcal{L}}(-|P_1(a_i) \wedge \neg P_1(a_j))$ but, as already discussed in the previous point, $c_0^{\mathcal{L}}(P_1(a_i) \wedge \neg P_1(a_j)) = 0$.

Similarly, we can argue that also in the polyadic setting, Theorem 1.2.4 isn't a good result as hoped.

Hence, if our formulation of degrees of belief in terms of probabilities on $\text{Sen}(\mathcal{L})$ is right, we must conclude that INV is a too-strong principle.

One path that can be run is trying to define a subclass of automorphisms the invariance under which should give a rational principle. There are some tentative definitions that deserve more study; we will present briefly one of these, the *Permutation Invariance Principle (PIP)*.

Definition 1.3.1. An automorphism γ of $M(\mathcal{L})$ *permutes state formulas* if for any $n \in \mathbb{N}^+$ there is a permutation $f_{\gamma,n}$ of the set of state formulas over variables $x_1, \dots, x_n$ such that, for any state formula $\Theta(x_1, \dots, x_n)$ and for any constants $a_{i_1}, \dots, a_{i_n}$, we have

$$\gamma(\Theta(a_{i_1}, \dots, a_{i_n})) = f_{\gamma,n}(\Theta)\{x_1/a_{i_1}, \dots, x_n/a_{i_n}\}.$$

A probability w on $\text{Sen}(\mathcal{L})$ satisfies the *Permutation Invariance Principle* if for any γ automorphism of $M(\mathcal{L})$ that permutes state formulas, and for any $\varphi \in \text{Sen}(\mathcal{L})$,

$$w(\varphi) = w(\gamma(\varphi)).$$

This weakening of INV agrees with how we decided to present the topic of symmetry in this work. Indeed, at the beginning of the chapter, we preferred to stress out the rationality of some principles (Ex, Px, SN, Vx) using arguments based on the different formalizations available for a given problem and on the ignorance of an agent in the zero-knowledge condition.

This kind of argument[12], which seems really reliable to us, doesn't justify, as already noticed, the whole INV principle: for instance, the automorphism γ used in the proof of Theorem 1.2.3 seems different from the ones that could be born from reasons related to different formalizations.

It makes sense, then, to find other principles, not as strong as INV, that can imply the principles deduced by such arguments: PIP accomplishes this request, i.e. it is sufficient to deduce Ex, Px, SN, Vx and is strictly weaker than INV (see [23]). Furthermore, it can be proved that PIP is stronger than the ones in the list above: thus, the open question is whether PIP rules out other principles that come out from INV but that are not considered so rational.

Instead of focusing on a comprehensive symmetry principle that encompasses all the most accepted ones, we can try to detect other "notions" from which we can derive rationality: for instance, we want that an agent ignores *irrelevant knowledge* in forming beliefs regarding a fact. In the unary case, this will lead us to principles like the

- *Johnson's Sufficientness Postulate (JSP):*[13]
  the value
  $$w\big(\alpha_j(a_{n+1})|\bigwedge_{i=1}^{n}\alpha_{h_i}(a_i)\big)^{14}$$
  depends only on n and on the cardinality of the set $\{i \leq n \ : \ h_i = j\}$.

This principle can be stated in this way: knowing which atoms are satisfied by $a_1, \dots, a_n$, to decide which degree of belief should be assigned to $\alpha_j(a_{n+1})$, an agent should ignore all the information but n and the number of the constants among $a_1, \dots, a_n$ that satisfy the j-th atom.

The difference between *symmetry* and *irrelevance* isn't neat and some principles follow from both arguments: an example of this is Ax. We have already shown that INV implies it (and PIP does too, in the unary setting) and we will see in the next chapter that we can derive it also from JSP.

What was said before about irrelevance regards the unary context; for the polyadic case, the question is a bit more delicate and will lead us to another important open problem.

In the polyadic setting, we can't talk about atoms and we have difficulties already in generalizing the Ax principle in this context: in the following, we will try to rephrase it in order to extend it. The principle that comes out of this irrelevance argument plays an important role in one of the major actual questions on the subject.

Notice that when the probability w satisfies Ex and Ax, the value $w(\wedge_{i=1}^{m}\alpha_{h_i}(b_i))$ depends only on the multiset $\{m_1, \dots, m_{2^q}\}$, where

$$m_j := \{i \in \{1, \dots, 2^q\} \ : \ h_i = j\}.$$

Hence, in this case, the state description $\Theta = \wedge_{i=1}^{m}\alpha_{h_i}(b_i)$ induces a partition on the set $\{b_1, \dots, b_m\}$ depending on which constants are indistinguishable one from the other using

[12] They seem to be encompassed by another principle called *Conformity*: for further details see [20], Chapter 28.

[13] See [13] and [20] for further details.

[14] Using the convention that in the case in which $w(\psi) = 0$, $w(\varphi|\psi) = c$ means that $w(\varphi \wedge \psi) = cw(\psi)$.

the information given by Θ. The various m_i's are only the size of the equivalence classes that Θ creates. This motivates the following definition.

Definition 1.3.2. Given a state description $\Theta(b_1, \dots, b_m)$, we define the equivalence relation $\sim_\Theta$ on the set $\{b_1, \dots, b_m\}$, by putting

$$b_i \sim_\Theta b_j \text{ if and only if } \Theta(b_1, \dots, b_m) \wedge b_i = b_j \text{ is consistent.}^{15}$$

The *spectrum* of Θ, denoted by $\mathcal{S}(\Theta)$, is the multiset of the sizes of the $\sim_\Theta$-equivalence classes on $\{b_1, \dots, b_m\}$.

A probability w that satisfies Ex, satisfies the *Spectrum Exchangeability Principle (Sx)* if for any two state descriptions $\Theta(b_1, \dots, b_m)$ and $\Theta'(b_1, \dots, b_m)$ with the same spectrum, $w(\Theta) = w(\Theta')$.

In the unary case, Sx corresponds to Ax for probabilities that satisfy Ex.

Example 1.3.1. Given a language with a binary relational symbol R, then:

- if $\Theta(a_1, a_2, a_3)$ is the state description $\bigwedge_{i=1}^3 \bigwedge_{j=1}^3 R(a_i, a_j)$, then according to Θ the constants are all indistinguishable and then $\mathcal{S}(\Theta) = \{3\}$;
- if $\Theta(a_1, a_2, a_3, a_4)$ is the conjunction of the following atomic formulas

$$\begin{array}{cccc} R(a_1, a_1) & \neg R(a_1, a_2) & R(a_1, a_3) & R(a_1, a_4) \\ R(a_2, a_1) & \neg R(a_2, a_2) & R(a_2, a_3) & \neg R(a_2, a_4) \\ R(a_3, a_1) & \neg R(a_3, a_2) & R(a_3, a_3) & R(a_3, a_4) \\ R(a_4, a_1) & R(a_4, a_2) & R(a_4, a_3) & R(a_4, a_4) \end{array}$$

  then the equivalence classes induced by $\sim_\Theta$ are $\{a_1, a_3\}, \{a_2\}, \{a_4\}$, hence $\mathcal{S}(\Theta) = \{2, 1, 1\}$.

Sx can be regarded as an irrelevance principle: in evaluating $w(\Theta)$ nothing matters but the spectrum of the state formula. When we talk about symmetry, this becomes an open problem: does Sx derive from a symmetry principle? Is there a class of automorphisms the invariance under which corresponds to Sx?

By what is known now, Sx is strictly weaker than INV, since more probabilities satisfy Sx but only one that satisfies INV, and however it is strictly stronger than PIP. These properties give rise to hope for Sx as a principle that can capture rationality and encompass symmetry and irrelevance, but there is a lot of work still to do.

[15] Here *consistent* is with regard to the equality axioms.

Chapter 2

Relevance, Irrelevance and the Carnap's Continuum

In this chapter, we will focus on other principles that can be used to determine the rationality of a probability: namely we will study the Irrelevance on the form of the already mentioned *Johnson's Sufficientness Postulate (JSP)* and we will show a nice charachterization of the probabilities that satisfy Ex and JSP; we will show that these two properties will determine a family of probabilites called the *Carnap's continuum*.

2.1 Relevance

Maybe before talking about *Irrelevance*, we can discuss about what should be *relevant* for the agent in assigning beliefs.

The idea here is that in determining the chance of drawing a plain blue ball from the urn, it should be relevant the number of plain blue balls already drawn by the agent: the more frequently we (or our rational agent) have seen some event in the past the more frequently we expect to see it in the future.

There are many ways to capture this idea within the formalization used so far: Carnap in [1] states the following principle with this aim.

Definition 2.1.1. A probability w on $\mathrm{Sen}(\mathcal{L})$ satisfies the *Principle of Instantial Relevance (PIR)* if and only if for any $\varphi(a_1, \dots, a_n) \in \mathrm{Sen}(\mathcal{L})$ and atom $\alpha(x)$ of $\mathcal{L}$, we have that

$$w(\alpha(a_{n+2})|\alpha(a_{n+1}) \wedge \varphi(a_1, \dots, a_n)) \geq w(\alpha(a_{n+2})|\varphi(a_1, \dots, a_n)).$$

An informal description of the principle is the following: if the agent knows something $(\varphi(a_1, \dots, a_n))$ the fact that a_{n+1} satisfies α should enhance the chance for the agent to believes that also a_{n+2} satisfies α.

As already said, the notions of *symmetry, irrelevance and relevance* are not so neatly distincted and the next theorem moves in this direction.

Theorem 2.1.1. *Ex implies PIR.*

Proof. Let w be a probability on $\mathrm{Sen}(\mathcal{L})$ that satisfies Ex. Suppose that the atoms of the language are $\alpha_1(x), \dots, \alpha_{2^q}(x)$ and that $a = w(\varphi(a_1, \dots, a_n))$.

Using Corollary 1.2.2, denoting with μ De Finetti prior of w, we have that

$$a = w(\varphi(a_1, \ldots, a_n)) = \int_{\mathbb{D}_{2^q}} w_{\vec{x}}(\varphi(a_1, \ldots, a_n))\ d\mu(\vec{x}),$$

$$w(\alpha_1(a_{n+1}) \wedge \varphi(a_1, \ldots, a_n)) = \int_{\mathbb{D}_{2^q}} x_1 w_{\vec{x}}(\varphi(a_1, \ldots, a_n))\ d\mu(\vec{x}),$$

$$w(\alpha_1(a_{n+2}) \wedge \alpha_1(a_{n+1}) \wedge \varphi(a_1, \ldots, a_n)) \overset{*}{=} \int_{\mathbb{D}_{2^q}} x_1^2 w_{\vec{x}}(\varphi(a_1, \ldots, a_n))\ d\mu(\vec{x}).$$

Therefore, w satisfies PIR if and only if

$$\Big(\int_{\mathbb{D}_{2^q}} w_{\vec{x}}(\varphi(a_1, \ldots, a_n))\ d\mu(\vec{x})\Big) \cdot \Big(\int_{\mathbb{D}_{2^q}} x_1^2 w_{\vec{x}}(\varphi(a_1, \ldots, a_n))\ d\mu(\vec{x})\Big) \geq \Big(\int_{\mathbb{D}_{2^q}} x_1 w_{\vec{x}}(\varphi(a_1, \ldots, a_n))\ d\mu(\vec{x})\Big)^2$$

The thesis follows from the Cauchy-Schwarz disequality. □

A similar argument shows that Ex actually implies the *Extended Principle of Instantial Relevance (EPIR):* for any $\varphi(a_1, \ldots, a_n) \in \mathrm{Sen}(\mathcal{L})$ and for any $\theta(x)$ $\mathcal{L}$-formula with x as the only free-variable,

$$w(\theta(a_{n+2}) | \theta(a_{n+1}) \wedge \varphi(a_1, \ldots, a_n)) \geq w(\theta(a_{n+2}) | \varphi(a_1, \ldots, a_n)).$$

2.2 Irrelevance and Carnap's Continuum

While in the previous section we were interested in what should be relevant for the agent in assigning beliefs, now we will focus on the *irrelevant knowledge*. In this perspective, we recall the *Johnson's Sufficientness Postulate (JSP)*: the value

$$w\big(\alpha_j(a_{n+1}) | \bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)\big)^1$$

depends only on n and on the cardinality of the set $\{i \leq n\ :\ h_i = j\}$.

Actually this principle, with the already discussed and agreed by all Ex, determines a family of probabilities depending on a parameter λ as specified by the following definition.

Definition 2.2.1. For a unary language $\mathcal{L} = \{R_1, \ldots, R_q\}$, for any $0 \leq \lambda \leq \infty$ the probability function $c_\lambda^{\mathcal{L}}$ is the only one such that

$$c_\lambda^{\mathcal{L}}\big(\alpha_j(a_{n+1}) | \bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)\big) = \frac{m_j + \lambda 2^{-q}}{n + \lambda}, \tag{2.1}$$

where $m_j = |\{i \leq n\ :\ h_i = j\}|$ and we use the convention that $2^{-q} \cdot \infty/\infty = 2^{-q} \cdot 0/0 = 2^{-q}$.

[1]Using the convention that in the case in which $w(\psi) = 0$, $w(\varphi | \psi) = c$ means that $w(\varphi \wedge \psi) = cw(\psi)$.

Proposition 2.2.1. *The definition above is well-posed: indeed, it defines a unique probability $c_\lambda^{\mathcal{L}}$ for any $\lambda \in [0, \infty]$.*

Proof. In the case $\lambda = 0$, we have already encountered the probability $c_0^{\mathcal{L}}$: we will show that the definition above agrees with the one given in Chapter 1:

$$c_0^{\mathcal{L}}\big(\bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)\big) = \prod_{j=1}^{n} c_0^{\mathcal{L}}\big(\alpha_{h_j}(a_j) | \bigwedge_{i=1}^{j-1} \alpha_{h_i}(a_i)\big); \tag{2.2}$$

- if all the h_i's are equal, then

  $$c_0^{\mathcal{L}}\big(\alpha_{h_j}(a_j) | \bigwedge_{i=1}^{j-1} \alpha_{h_i}(a_i)\big) = \frac{(j-1)}{(j-1)} = 1,$$

  for $j > 1$, but for $j = 1$ we have

  $$c_0^{\mathcal{L}}\big(\alpha_{h_1}(a_1) | \bigwedge_{i=1}^{0} \alpha_{h_i}(a_i)\big) = \frac{0 + 0 \cdot 2^{-q}}{0+0} = 2^{-q},$$

  hence the thesis;

- otherwise, if not all the h_i's are equal, there exists a $k > 1$ such that $h_1 = \cdots = h_{k-1}$ and $h_k \neq h_1$. Hence,

  $$c_0^{\mathcal{L}}\big(\alpha_{h_k} | \bigwedge_{i=1}^{k-1} \alpha_{h_i}(a_i)\big) = \frac{0 + 0 \cdot 2^{-q}}{k-1} = 0$$

  and

  $$c_0^{\mathcal{L}}\big(\bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)\big) = 0.$$

For $\lambda > 0$, by Equation (2.1), we have that

$$\begin{aligned} c_\lambda^{\mathcal{L}}\big(\bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)\big) &= \prod_{j=1}^{n} c_\lambda^{\mathcal{L}}\big(\alpha_{h_j}(a_j) | \bigwedge_{i=1}^{j-1} \alpha_{h_i}(a_i)\big) \\ &= \prod_{j=1}^{n} \Big(\frac{r_j + \lambda 2^{-q}}{j - 1 + \lambda}\Big) \\ &= \frac{\prod_{k=1}^{2^q} \prod_{j=0}^{m_k - 1} (j + \lambda 2^{-q})}{\prod_{j=0}^{n-1} (j + \lambda)}, \end{aligned} \tag{2.3}$$

where r_j is the number of times that h_j occurs among the $h_1, \ldots, h_{j-1}$ and m_k is the number of times k occurs among $h_1, \ldots, h_n$. Then the conditions established at page 17 are satisfied since:

- $c_\lambda^{\mathcal{L}}(\wedge_{i=1}^{n} \alpha_{h_i}(a_i)) \geq 0$ because it's a product of positive entities;
- $c_\lambda^{\mathcal{L}}(\top) = c_\lambda^{\mathcal{L}}(\wedge_{i=1}^{0} \alpha_{h_i}(a_i)) = 1$ since the empty product is 1;

- given $r \geq n$ it should be that

$$\sum_{f:\{n+1,\ldots,r\}\to\{1,\ldots,2^q\}} c_\lambda^{\mathcal{L}}\Big(\bigwedge_{i=1}^{n} \alpha_{h_i}(a_i) \wedge \bigwedge_{j=n+1}^{r} \alpha_{f_j}(a_i)\Big) = c_\lambda^{\mathcal{L}}\Big(\bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)\Big).$$

We will show that this holds when $r = n+1$ even if the same argument can be generalized for a generic $r \geq n$.

Denoting m_k the times k occurs in $h_1, \ldots, h_n$ and t_{kf} the times k occurs in $h_1, \ldots, h_n, f_{n+1}$

$$\sum_{f:\{n+1\}\to\{1,\ldots,2^q\}} \frac{\prod_{k=1}^{2^q}\prod_{j=0}^{t_{kf}-1}(j+\lambda 2^{-q})}{\prod_{j=0}^{n-1}(j+\lambda)}$$
$$= \frac{\prod_{k=1}^{2^q}\prod_{j=0}^{m_k-1}(j+\lambda 2^{-q})}{\prod_{j=0}^{n-1}(j+\lambda)} \cdot \frac{(m_1+\lambda 2^{-q})}{n+\lambda} + \cdots + \frac{\prod_{k=1}^{2^q}\prod_{j=0}^{m_k-1}(j+\lambda 2^{-q})}{\prod_{j=0}^{n-1}(j+\lambda)} \cdot \frac{(m_{2^q}+\lambda 2^{-q})}{n+\lambda}$$
$$= \frac{\prod_{k=1}^{2^q}\prod_{j=0}^{m_k-1}(j+\lambda 2^{-q})}{\prod_{j=0}^{n-1}(j+\lambda)} \cdot \frac{(m_1+\cdots+m_{2^q}+2^q\cdot\lambda 2^{-q})}{n+\lambda}$$
$$= \frac{\prod_{k=1}^{2^q}\prod_{j=0}^{m_k-1}(j+\lambda 2^{-q})}{\prod_{j=0}^{n-1}(j+\lambda)} \cdot \frac{n+\lambda}{n+\lambda} = \frac{\prod_{k=1}^{2^q}\prod_{j=0}^{m_k-1}(j+\lambda 2^{-q})}{\prod_{j=0}^{n-1}(j+\lambda)}.$$

□

It's easy to see that the first symmetry principles encountered in the previous chapter are satisfies by the $c_\lambda^{\mathcal{L}}$, as the following proposition shows.

Proposition 2.2.2. *The Carnap's Continuum $c_\lambda^{\mathcal{L}}$ satisfy Ex, Px, SN and Ax.*

Proof. We have already shown that INV implies all the principles listed above, so when $\lambda = 0$ the proposition has been already proved.

For $\lambda > 0$, from Equation 2.3, we can see that $c_\lambda^{\mathcal{L}}$ doesn't depend on the a_i's, hence it satisfies Ex; since the second term of Equation 2.3 depends only on the set $\{m_1, \ldots, m_{2^q}\}$, we have that $c_\lambda^{\mathcal{L}}$ satisfies Ax.

Replacing a relational symbol with another or a relational symbol with its negation in a conjunction of atoms is equivalent to apply a specific permutation of atoms: hence, Ax implies SN and Px and in particular all the $c_\lambda^{\mathcal{L}}$'s satisfy SN and Px. □

As already announced, this family comes out by investigating a little further the JSP principle and we will need the next theorem to understand better what it means.

Theorem 2.2.1. *Suppose that $\mathcal{L}$ has at least two realtional symbols (i.e. $q \geq 2$)[2]; a probability w satisfies Ex and JSP if and only if is of the form $c_\lambda^{\mathcal{L}}$ for some $\lambda \in [0, \infty]$.*

Proof. Clearly, by the definitions, the Carnap's Continuum satisfies JSP and we have already shown that also Ex holds for them.

For the converse, the first thing to notice is that JSP implies Ax. First, assume that there is a state description $\wedge_{i=1}^{n}\alpha_{h_i}(a_i)$ to which w gives the value 0; we can without loss of generality suppose that n is the minimal with this property. Then

[2]When $q = 1$ it can be shown that Ax, SN, and JSP are all equivalent principle, hence there are more probabilities that satisfy JSP.

- $n > 1$, since we have that

$$1 = w\big(\bigvee_{i=1}^{2^q} \alpha_i(a_i)\big) = \sum_{i=1}^{2^q} w(\alpha_i(a_i))$$

  and for Ax, to any atom is assigned the same value, i.e. 2^{-q};

- if two atoms in $\wedge_{i=1}^{n}\alpha_{h_i}(a_i)$ are equal, say that $h_1 = h_2$, then by PIR (and then by Ex, by Theorem 2.1.1), we have that

$$0 = w\big(\alpha_{h_1}(a_1)|\bigwedge_{i=2}^{n} \alpha_{h_i}(a_i)\big) \geq w\big(\alpha_{h_1}(a_1)|\bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)\big),$$

  contradicting the minimality of n.

Hence all the h_i's must be different and by JSP, we have that

$$0 = w\big(\alpha_{h_1}(a_1)|\bigwedge_{i=2}^{n} \alpha_{h_i}(a_i)\big) = w\big(\alpha_1(a_1)|\bigwedge_{i=2}^{n} \alpha_2(a_i)\big),$$

therefore $w(\alpha_1(a_1) \wedge \wedge_{i=2}^{n}\alpha_2(a_i)) = 0$.

For what said before, all the h_i's that appears in the minimal state description to which w assigns the null value must be distinct, hence $n = 2$ and

$$w(\alpha_1(a_1) \wedge \alpha_2(a_2)) = 0;$$

by Ax, this means that for all n,

$$w\big(\bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)\big) = 0$$

if the h_i's are not all equal, and this directly leads to $w = c_0^{\mathcal{L}}$.

Suppose now to have a probability w that satisfies JSP and assume that is non-zero in the state descriptions. Then, by JSP the quantity $w(\alpha_j(a_{n+1})|\wedge_{i=1}^{n}\alpha_{h_i}(a_i))$ depends only on n and r, i.e. the times j occurs in $h_1, \ldots, h_n$, hence this value can be denoted by $g(r, n)$.

$$\begin{aligned} 1 &= w\big(\bigvee_{i=1}^{2^q} \alpha_i(a_2)|\alpha_j(a_1)\big) = \sum_{i=1}^{2^q} w(\alpha_i(a_2)|\alpha_j(a_1)) \\ &= g(1,1) + (2^q - 1)g(0,1). \end{aligned} \tag{2.4}$$

Since w satisfies Ex, it satisfies also PIR by Theorem 2.1.1: hence,

$$g(1,1) = w(\alpha_j(a_2)|\alpha_j(a_1)) \geq w(\alpha_j(a_2)) = g(0,0).$$

By Ax, $1 \geq g(1,1) \geq 2^{-q}$, hence $g(1,1)$ can be written as a convex combination of 1 and 2^{-q}, i.e.

$$g(1,1) = \frac{1}{1+\lambda} + \frac{\lambda}{1+\lambda}2^{-q} = \frac{1+2^{-q}\lambda}{1+\lambda},$$

for some $\lambda \in [0, \infty]$. Hence, by Equation 2.4, we have that

$$g(0,1) = \frac{2^{-q}\lambda}{1+\lambda}.$$

Since w is non-zero in all the state descriptions, $g(0,1) > 0$, hence $\lambda > 0$.

We can show now that these values forces for any $r, n \in \mathbb{N}$ with $r \leq n$ that

$$g(r,n) = \frac{r + \lambda 2^{-q}}{n+\lambda} : \tag{2.5}$$

- for $n = 0, 1$ we have already proved the thesis with all the values $r \leq n$ possible;
- assume that Equation 2.5 holds for n and all the $r \leq n$. Let's take now r, s such that $r + s = n + 1$ and two atom indexs $k, m \in \{1, \ldots, 2^q\}$. We have that

$$\begin{aligned} 1 &= w\big(\bigvee_{h=1}^{2^q} \alpha_h(a_{n+1}) | \bigwedge_{i=1}^{r} \alpha_m(a_i) \wedge \bigwedge_{i=r+1}^{n+1} \alpha_k(a_i)\big) \\ &= \sum_{h=1}^{2^q} w\big(\alpha_h(a_{n+1}) | \bigwedge_{i=1}^{r} \alpha_m(a_i) \wedge \bigwedge_{i=r+1}^{n+1} \alpha_k(a_i)\big) \\ &= g(r, n+1) + g(s, n+1) + (2^q - 2) g(0, n+1). \end{aligned} \tag{2.6}$$

Using the notation $\alpha_{h_1}\alpha_{h_2}\ldots\alpha_{h_n}$ to denote $\wedge_{i=1}^{n}\alpha_{h_i}$ and taking r, s, t such that $r + s + t = n$ and distinct atom indexs $m, j, k \in \{1, \ldots, 2^q\}$[3], we have that

$$\begin{aligned} w(\alpha_m|\alpha_j\alpha_m^r\alpha_j^s\alpha_k^t) \cdot w(\alpha_j|\alpha_m^r\alpha_j^s\alpha_k^t) &= w(\alpha_m\alpha_j|\alpha_m^r\alpha_j^s\alpha_k^t) \\ w(\alpha_j|\alpha_m\alpha_m^r\alpha_j^s\alpha_k^t) \cdot w(\alpha_m|\alpha_m^r\alpha_j^s\alpha_k^t), \end{aligned}$$

hence

$$g(r, n+1) \cdot g(s, n) = g(s, n+1) \cdot g(r, n). \tag{2.7}$$

Taking $s = 0$, we have for the inductive hypothesis

$$g(r,n+1) = \frac{g(r,n)}{g(0,n)} g(0,n+1) = \frac{r + \lambda 2^{-q}}{n+\lambda} \frac{n+\lambda}{\lambda 2^{-q}} g(0,n+1) = (r\lambda^{-1}2^q + 1) g(0,n+1). \tag{2.8}$$

Taking $r = 1, s = n$ in Equation 2.6 and replacing the outcome in $g(1, n+1)$ and $g(n, n+1)$ in Equation 2.8, gives

$$\begin{aligned} 1 &\overset{2.6}{=} g(1, n+1) + g(n, n+1) + (2^q - 2) g(0, n+1) \\ &\overset{2.8}{=} (\lambda^{-1}2^q + 1) g(0, n+1) + (n\lambda^{-1}2^q + 1) g(0, n+1) + (2^q - 2) g(0, n+1), \end{aligned}$$

from which

$$g(0, n+1) = \frac{\lambda 2^{-q}}{n + 1 + \lambda}.$$

The thesis for a generic $r \leq n + 1$ follows by replacing using the equation above $g(0, n+1)$ in Equation 2.8.

[3]This is possible since $q \geq 2$

□

The $c_\lambda^{\mathcal{L}}$'s have been as a good family to represent the beliefs of a rational agent: as already proved, they satisfisfy the basic symmetry principles (Ex, SN, Px, Ax) and JSP. Furthermore, they satisfy also the *Reichenbach Axiom (RA)* a famous principle in the applicative counterpart of this theory.

Definition 2.2.2. A probability w on Sen($\mathcal{L}$) satisfies the *Reichenbach Axiom (RA)* if and only if for any function $f : \mathbb{N}^+ \to \{1, \ldots, 2^q\}$ and for any $\mathcal{L}$-atom $\alpha_j(x)$, we have that

$$\lim_{n\to+\infty}\big(w\big(\alpha_j(a_{n+1})|\bigwedge_{i=1}^{n}\alpha_{h_i}(a_i)\big) - \frac{u_j(n)}{n}\big) = 0,$$

where $u_j(n) = |\{i \ : \ 1 \le i \le n \text{ and } h_i = j\}|$.

Obviously, for the definition of $c_\lambda^{\mathcal{L}}$, we have that

$$\begin{aligned} c_\lambda^{\mathcal{L}}\big(\alpha_j(a_{n+1})|\bigwedge_{i=1}^{n}\alpha_{h_i}(a_i)\big) - \frac{u_j(n)}{n} &= \frac{u_j(n)+\lambda 2^{-q}}{n+\lambda} - \frac{u_j(n)}{n} \\ &= \frac{n\lambda 2^{-q} - \lambda u_j(n)}{n(n+\lambda)}, \end{aligned}$$

and the limit of this as n approaches to $+\infty$ is

$$\lim_{n\to+\infty}\frac{n\lambda 2^{-q} - \lambda u_j(n)}{n(n+\lambda)} = \lim_{n\to+\infty}\frac{\lambda 2^{-q}}{n+\lambda} - \lim_{n\to+\infty}\frac{\lambda u_j(n)}{n(n+\lambda)} = 0.$$

However, also this family has its faults: in particular, one is that no amount of evidence of the form $\wedge_{i=1}^{n}\varphi(a_i)$ can lead the agent whose beliefs are according to $c_\lambda^{\mathcal{L}}$ to give the value 1 to $\forall x\varphi(x)$ when $\lambda > 0$. Indeed, also when $\varphi(x)$ is a $\mathcal{L}$-atom, say $\alpha(x)$, we have that

$$\begin{aligned} c_\lambda^{\mathcal{L}}\big(\forall x\alpha(x)|\bigwedge_{i=1}^{n}\alpha(a_i)\big) &= \lim_{m\to+\infty} c_\lambda^{\mathcal{L}}\big(\bigwedge_{j=1}^{m}\alpha(a_j)|\bigwedge_{i=1}^{n}\alpha(a_i)\big) \\ &= \lim_{m\to+\infty} c_\lambda^{\mathcal{L}}\big(\alpha(a_m)|\bigwedge_{i=1}^{m-1}\alpha(a_i)\big)\cdot c_\lambda^{\mathcal{L}}\big(\alpha(a_{m-1})|\bigwedge_{i=1}^{m-2}\alpha(a_i)\big)\cdots c_\lambda^{\mathcal{L}}\big(\alpha(a_{n+1}|\bigwedge_{i=1}^{n}\alpha(a_i)\big) \\ &= \lim_{m\to+\infty}\frac{m-1+\lambda 2^{-q}}{m-1+\lambda}\cdot\frac{m-2+\lambda 2^{-q}}{m-2+\lambda}\cdots\frac{n+\lambda 2^{-q}}{n+\lambda} \\ &= \lim_{m\to+\infty}\big(1-\frac{\lambda(1-2^{-q})}{m-1+\lambda}\big)\cdot\big(1-\frac{\lambda(1-2^{-q})}{m-2+\lambda}\big)\cdots\big(1-\frac{\lambda(1-2^{-q})}{n+\lambda}\big) \\ &\le \lim_{m\to+\infty}\big(1-\frac{\lambda(1-2^{-q})}{m-1+\lambda}\big)^{m-n} = \lim_{m\to+\infty} e^{-\lambda(1-2^{-q})} < 1. \end{aligned}$$

We have already argued the flaws of $c_0^{\mathcal{L}}$, but we can notice that in this case, this probability is able to capture the validity of Π_1-sentences.

$$c_0^{\mathcal{L}}\big(\forall x\alpha(x)|\bigwedge_{i=1}^{n}\alpha(a_i)\big) = \lim_{m\to+\infty} c_0^{\mathcal{L}}\big(\bigwedge_{j=1}^{m}\alpha(a_j)|\bigwedge_{i=1}^{n}\alpha(a_i)\big) = 1.$$

However, Carnap himself and other scholars argued that is sufficient to ask a probability w that an evidence of the form $\wedge_{i=1}^{n}\varphi(a_i)$ leads the agent to believe in $\varphi(a_{n+1})$ assigning to it the value 1, instead of recongnising the validity of $\forall x\varphi(x)$ when n approaches to ∞: this property holds for the $c_\lambda^{\mathcal{L}}$'s since it can be easily derived from the Reichenbach Axiom.

Another more serious problem of the $c_\lambda^{\mathcal{L}}$ is linked to Example 1.2.1. Since the $c_\lambda^{\mathcal{L}}$'s satisfy Ex, they can be written as in Theorem 1.2.1: indeed, we can show that

$$c_\lambda^{\mathcal{L}} = k_q \int_{\mathbb{D}_{2^q}} w_{\vec{x}} \prod_{i=1}^{2^q} x_i^{\lambda 2^{-q}-1}\ d\mu_q(\vec{x}),$$

where q is, as usual, the number of relational symbols of the language, k_q is a normalization constant, and μ_q is the Lebsegue measure on $\mathbb{D}_{2^q}$.

In the Example 1.2.1, we chose a probability (the Lebsegue's one in $[0,1]$) that assigned to all the $w_{\vec{x}}$'s the same value and we get for the unary language considered the probability c_2; in general, using the equation above, we can see that for a language $\mathcal{L}$ with q relational symbols, in order to give the same weight to all the $w_{\vec{x}}$'s, we should take $\lambda = 2^q$.

We can show that if we remove from the language $\mathcal{L}$ a relational symbol, getting the language $\mathcal{L}^-$, then $c_\lambda^{\mathcal{L}}\big|_{\mathrm{Sen}(\mathcal{L}^-)} = c_\lambda^{\mathcal{L}^-}$, therefore, for $\lambda = 2^q$, we have

$$c_{2^q}^{\mathcal{L}}\big|_{\mathrm{Sen}(\mathcal{L}^-)} = k_{q-1} \int_{\mathbb{D}_{2^{q-1}}} w_{\vec{x}} \prod_{i=1}^{2^q-1} x_i\ d\mu_{q-1}(\vec{x}).$$

Hence, if we consider rational to assign the same weight to all the $w_{\vec{x}}$'s, we have to pay attention to the language considered: when $\mathcal{L}$ has q relational symbols, in $c_{2^q}^{\mathcal{L}}$ the $w_{\vec{x}}$'s are equally wighted, but when we restrict to a sublanguage, this doens't happen anymore if we mantain the same probability.

This attention for the languages is motivated by the fact that we want a consistency between the different languages and the respective probabilities that can be used; indeed, it would appear unreasonable to assume from the start that $\mathcal{L}$ was all the language there could ever be; maybe we discover that there are some properties that needs further relational symbols, or maybe we discover that some relational symbols are superfluous since equivalent, for example, to a combination of the others.

Hence, to sum up, the family of the $c_\lambda^{\mathcal{L}}$'s seems to satisfy a lot of basic principles (Ex, Px, SN, Ax, JSP) and it can be argued that these probabilities achieve most of our requirements about a rational agent's beliefs. If we want to delve into the choice of the specific λ, then we have some issues to consider: the case $\lambda = 0$ is dealt when we talk about symmetry in Chapter 1 and it doesn't seem to be a good choice; the case $\lambda = 2^q$ seems reasonable even if we lack still a good reason to choose this value: in particular, the motivation rooted to the fact of giving the same weight to the probabilities involved in its De Finetti's representation depends too much on the language we are working with.

Bibliography

[1] Carnap R., *Logical Foundations of Probability*, University of Chicago Press, Routledge, Kegan Paul, London, 1950.

[2] Christiano P., *Non-Omniscience, Probabilistic Inference, and Metamathematic*, MIRI, 2014.

[3] Cohn D. L., *Measure Theory*, Birkäuser Advanced Test, Second edition, Springer, 2013.

[4] Cutland N.J., *Loeb Measure Theory*, Developments in Nonstandard Mathematics (with Oliveira F., Neves V., and Sousa-Pinto J.), Pitman Research Notes in Mathematics Series, Vol. 336, Longman CRC Press, pp. 151-177.

[5] Demski A., *Logical Prior Probability*, Artificial General Intelligence, Springer, 2012.

[6] Fagin R., *Probabilities on Finite Models*. The Journal of Symbolic Logic, Vol. 41, No. 1, pp. 50-58, 1976.

[7] Gaifman H. and Snir M., *Probabilities Over Rich Languages*. The Journal of Symbolic Logic, Vol. 47, No. 3, pp. 495-548, 1982.

[8] Garrabrant S., Benson-Tilsen T., Critch A., Soares N., and Taylor J., *Logical Induction*, Machine Intelligence Research Institution, 2016.

[9] Goldblatt R., *Lectures on the Hyperreals: An Introduction to Nonstandard Analysis.* Graduate Texts in Mathematics, Springer, 1998.

[10] Halpern J., *Reasoning about Uncertainty*, MIT Press, 2003.

[11] Howarth E., Paris J., *Pure Inductive Logic with Functions*, The Journal of Symbolic Logic, 2019.

[12] Hutter M., Lloyd J.W., Ng K.S. and Uther W.T.B., *Probabilities on Sentences in an Expressive Logic*. Presented at the Fifth Workshop on Combining Probability and Logic, 2012.

[13] Johnson W. E., *Probability: The deductive and inductive problems*, Mind, Vol. 41, 1932.

[14] Jonas H., *Das Prinzip Verantwortung - The imperative of Responsibility: In Search of an Ethics for the Technological Age*, English version published by Einaudi, 1979.

[15] Krauss P. H., *Representation of Symmetric Probability Models*. The Journal of Symbolic Logic, Vol. 34, No. 2, pp. 183-193, 1969.

[16] Kunen K., *Set Theory: An Introduction to Independence Proofs*, Elsevier Science Publisheres B. V., 1980.

[17] Monk D.J., Bonnet R., *Handbook of Boolean Algebras*, Vol. 1, Elsevier, 1989.

[18] Moses Y., Halpern J., Fagin R., Vardi M., *Reasoning about Knowledge*, MIT Press, 1989.

[19] Paris J., Vencovská A., *On the Strongest Principles of Rational Belief Assignments*, Personal communication, draft under publication, 2024.

[20] Paris J., Vencovská A., *Pure Inductive Logic*. Perspectives in Logic, Cambridge University Press, 2015.

[21] Paris J., Vencovská A., *Six Problems in Pure Inductive Logic*, Journal of Philosophical Logic, Vol. 48, No. 4, pp. 731-747, 2019.

[22] Paris J., Vencovská A., *Symmetry's End?*, Erkenntnis, Vol. 74, No. 1, pp. 53-67, 2011.

[23] Paris, J.B., Vencovská, A., *Symmetry Principles in Pure Inductive Logic*, in *Logica 2022*, College Publications, U.K., Ed, Igor Sedl´ar, 2022.

[24] Ramsey F. P., *Truth and Probability* (1926). R. B. Braithwaite (Ed.), The Foundations of Mathematics and other Logical Essays, Routledge and Kegan Paul, London, 1931, pp- 156-198.

[25] Soares N., Fallenstein B., *Agent Foundations for Aligning Machine Intelligence with Human Interests: A Technical Research Agenda*, MIRI, 2016.

[26] Master's Thesis of Michele Mignani, *Inductive Logic*, 2024.