

## DEEP COMPUTATIONS AND NIP

2 EDUARDO DUEÑEZ<sup>1</sup> JOSÉ IOVINO<sup>1</sup> TONATIUH MATOS-WIEDERHOLD<sup>2</sup>  
3 LUCIANO SALVETTI<sup>2</sup> FRANKLIN D. TALL<sup>2</sup>

<sup>1</sup>Department of Mathematics, University of Texas at San Antonio  
<sup>2</sup>Department of Mathematics, University of Toronto

**ABSTRACT.** This paper revisits and extends a bridge between functional analysis and model theory, emphasizing its relevance to the theoretical foundations of machine learning. We show that the compactness behavior of families of Baire class 1 functions mirrors the learnability conditions in the sense of *Probably Approximately Correct* (PAC) learning, and that the failure of compactness corresponds to the presence of infinite *Vapnik-Chervonenkis* (VC) dimension. From this perspective, Rosenthal compacta emerge as the natural topological counterpart of PAC-learnable concept classes, while NIP vs. IP structures capture the precise boundary between analytical regularity and combinatorial intractability. These parallels suggest a unified framework linking compactness, definability, and learnability, exemplifying how the topology of function spaces encodes the algorithmic and epistemic limits of prediction.

## 1. INTRODUCTION

In this paper we study limit behavior of real-valued computations as the value of certain parameters of the computation model tend towards infinity or zero, e.g., the depth of a neural network tending to infinity or the time interval between layers of the network tending toward zero. Recently, particular cases of this situation have attracted considerable attention in the machine learning literature (e.g., neural ODE’s [CRBD] or deep equilibrium models [BKK]). In this paper, we combine ideas of topology and model theory to study these limit phenomena from a more general viewpoint. Informed by model theory, to each computation in a given computation model, we associate a continuous real-valued function called the *type* of the computation. This allows us to view computations in a given computational model as elements of a space of real-valued functions, called the *space of types* of the model, and thereby to utilize the vast theory of topology of function spaces, known as  $C_p$ -theory, to obtain results about complexity of topological limits of computations. As we indicate next, recent classification results for topological spaces of functions provide and elegant and powerful machinery to classify computations according to their level “tameness” or “wildness”, with the former corresponding to polynomial approximability and the latter to or exponential approximability. The viewpoint of spaces of types, which we borrow from model theory, thus becomes a “Rosetta stone” that allows us to interconnect various classification programs: In topology, the classification of Rosenthal compacta pioneered by Todorčević [Tod99]; in logic, the classification of theories due Shelah [She90]; and in statistical learning, the notion PAC learning and VC dimension pioneered by Vapkins and Chervonenkis [?, VC15].

In a previous paper [ADIW24], we introduced the concept of limits of computations, which we called *ultracomputations* (given they arise as ultrafilter limits of

<sup>32</sup> standard computations) and *deep computations* (following usage in machine learning [BKK]). There is a technical difference between both, but in this paper, to <sup>33</sup> simplify the nomenclature, we will ignore the difference and use only the term <sup>34</sup> “deep computation”.

<sup>35</sup> In [ADIW24], we investigated deep computations (or ultracomputations) that <sup>36</sup> are (real-valued) continuous functions. Under our model-theoretic Rosetta stone, <sup>37</sup> polynomial approximability in the sense of computation becomes identified with <sup>38</sup> the notion of continuous extendability in the sense of topology, and to the notion <sup>39</sup> of *stability* in the sense of model theory.

<sup>40</sup> In this paper, we follow the general approach, i.e., we investigate ultracomputations <sup>41</sup> are pointwise limits of continuous functions. In topology, real-valued functions <sup>42</sup> that arise as the pointwise of a sequence of continuous are called *Baire class 1* <sup>43</sup> functions, or *Baire-1* for short; they form a step above simple continuity in the <sup>44</sup> hierarchy of functions studied in real analysis (Baire class 0 functions being continuous <sup>45</sup> functions). Intuitively, Baire-1 functions represent functions with “controlled” <sup>46</sup> discontinuities, and they are therefore crucial in topology and set theory.

<sup>47</sup> In the first paper, which focused on continuous deep computations, we invoked a <sup>48</sup> classical result of Grothendieck from late 50s [Gro52] to obtain a new polynomial- <sup>49</sup> vs-exponential dichotomy for deep computations. In this paper, which focuses on <sup>50</sup> general Baire-1 computations, we invoke a celebrated result of Todorčević from the <sup>51</sup> late 90s, for Rosenthal compacta [Tod99], to obtain a new trichotomy of general <sup>52</sup> deep computations. Through the aforementioned Rosetta stone, Rosenthal compacta <sup>53</sup> in topology correspond to the important concept of No Independence Property <sup>54</sup> (known as “NIP”) in model theory [She71, She90], and to the concept of Prob- <sup>55</sup> ably Approximately Correct learning (known as “PAC learnability”) in statistical <sup>56</sup> learning theory [?, VC15]. We then go beyond Todorčević’s trichotomy, and invoke <sup>57</sup> a more recent heptachotomy for minimal families from the early 2000s [ADK08].

<sup>58</sup> We believe that the results presented here show practitioners of computation, or <sup>59</sup> topology or, or model theory, how classification invariants in their field translate <sup>60</sup> into classification invariants of other fields. However, in the interest of accessibility, <sup>61</sup> we do not assume previous familiarity with high level topology or model theory, <sup>62</sup> or computing. The only technical prerequisite of the paper is undergraduate-level <sup>63</sup> topology. The necessary topological preliminaries are included in section 3.

<sup>64</sup> Throughout the paper, we focus on classical computation; however, the results <sup>65</sup> presented here can be extended, using contemporary model-theoretic machinery, to <sup>66</sup> quantum computation and open quantum systems. This extension will be addressed <sup>67</sup> in a forthcoming paper.

## <sup>69</sup> 2. MOTIVATION

<sup>70</sup> Suppose that  $A$  is a subset of the real line  $\mathbb{R}$  and that  $\overline{A}$  is its *closure*. It is a <sup>71</sup> well-known fact that any point of closure of  $A$ , say  $x \in \overline{A}$ , can be *approximated* <sup>72</sup> by points inside of  $A$ , in the sense that a sequence  $\{x_n\}_{n \in \mathbb{N}} \subseteq A$  must exist with <sup>73</sup> the property that  $\lim_{n \rightarrow \infty} x_n = x$ . For most applications we wish to approximate <sup>74</sup> objects more complicated than points, such as functions.

<sup>75</sup> Suppose we wish to build a neural network that decides, given an 8 by 8 black- <sup>76</sup> and-white image of a hand-written scribble, what single decimal digit the scrib- <sup>77</sup> ble represents. Maybe there exists  $f$ , a function representing an optimal solution <sup>78</sup> to this classifier. Thus if  $X$  is the set of all (possible) images, then for  $I \in X$ ,

<sup>79</sup>  $f(I) \in \{0, 1, 2, \dots, 9\}$  is the “best” (or “good enough” for whatever deployment is  
<sup>80</sup> needed) possible guess. Training the neural network involves approximating  $f$  until  
<sup>81</sup> its guesses are within an acceptable error range. In general,  $f$  might be a function  
<sup>82</sup> defined on a more complicated topological space  $X$ .

<sup>83</sup> Often computers’ viable operations are restricted (addition, subtraction, multi-  
<sup>84</sup> plication, division, etc.) and so we want to approximate a complicated function  
<sup>85</sup> using simple functions (like polynomials). The problem is that, in contrast with  
<sup>86</sup> mere points, functions in the closure of a set of functions need not be approximable  
<sup>87</sup> (meaning the pointwise limit of a sequence of functions) by functions in the set.

<sup>88</sup> Functions that are the pointwise limit of continuous functions are *Baire class 1*  
<sup>89</sup> *functions*, and the set of all of these is denoted by  $B_1(X)$ . Notice that these are  
<sup>90</sup> not necessarily continuous themselves! A set of Baire class 1 functions,  $A$ , will be  
<sup>91</sup> relatively compact if its closure consists of just Baire class 1 functions (we delay the  
<sup>92</sup> formal definition of *relatively compact* until Section 3, but the fact mentioned here  
<sup>93</sup> is sufficient). The Bourgain-Fremlin-Talagrand (BFT) theorem reveals a precise  
<sup>94</sup> correspondence between relative compactness in  $B_1(X)$  and the model-theoretic  
<sup>95</sup> notion of *Non-Independence Property* (NIP). This was realized by Pierre Simon in  
<sup>96</sup> [Sim15b].

<sup>97</sup> Simon’s insight was to view definable families of functions as sets of real-valued  
<sup>98</sup> functions on type spaces and to interpret relative compactness in  $B_1(X)$  as a form  
<sup>99</sup> of “tame behavior” under ultrafilter limits. From this perspective, NIP theories are  
<sup>100</sup> those whose definable families behave like relatively compact sets of Baire class 1  
<sup>101</sup> functions, avoiding the wild,  $\beta\mathbb{N}$ -like configurations that witness instability. This  
<sup>102</sup> observation opened a new bridge between analysis and logic: topological compact-  
<sup>103</sup> ness corresponds to the absence of combinatorial independence. Simon’s later de-  
<sup>104</sup> velopments connected these ideas to *Keisler measures* and *empirical averages*, al-  
<sup>105</sup> lowing tools from functional analysis to be used to study learnability and definable  
<sup>106</sup> types. This reinterpretation of model-theoretic tameness through the lens of the  
<sup>107</sup> BFT theorem has made NIP a central notion not only in stability theory but also  
<sup>108</sup> in contemporary connections with learning theory and ergodic analysis.

Historically, the notion of NIP arises from Shelah’s foundational work on the classification theory of models. In his seminal book *Unstable Theories* [She78], Shelah introduced the independence property as a key dividing line within unstable structures, identifying the class of *stable* theories inside those in which this property fails. Fix a first-order formula  $\varphi(x, y)$  in a language  $L$  and a model  $M$  of an  $L$ -theory  $T$ . We say that  $\varphi(x, y)$  has the *independence property (IP)* in  $M$  if there is a sequence  $(c_i)_{i \in \mathbb{N}} \subseteq M^{|x|}$  such that for every  $S \subseteq \mathbb{N}$  there is  $a_S \in M^{|y|}$  with

$$\forall i \in \mathbb{N}, \quad M \models \varphi(c_i, a_S) \iff i \in S.$$

<sup>109</sup> The formula  $\varphi(x, y)$  has the IP if it does so in some model  $M$ , and the formula  
<sup>110</sup> has the *non-independence property (NIP)* if it does not have the IP. The latter  
<sup>111</sup> notion of NIP generalizes stability by forbidding the full combinatorial indepen-  
<sup>112</sup> dence pattern while allowing certain controlled forms of instability. Thus, Simon’s  
<sup>113</sup> interpretation of the BFT theorem can be viewed as placing Shelah’s dividing line  
<sup>114</sup> into a topological-analytic framework, connecting the earliest notions of stability  
<sup>115</sup> to compactness phenomena in spaces of Baire class 1 functions.

<sup>116</sup> One of the most important innovations in Machine Learning is the mathemati-  
<sup>117</sup> cal notion, introduced by Turing Awardee Leslie Valiant in the 1980s, of ‘probably

118 approximately correct learning', or PAC-learning for short [BD19]. We give a stan-  
 119 dard but short overview of these concepts in the context that is relevant to this  
 120 work.

121 Consider the following important idea in data classification. Suppose that  $A$  is  
 122 a set and that  $\mathcal{C}$  is a collection of sets. We say that  $\mathcal{C}$  *shatters*  $A$  if every subset  
 123 of  $A$  is of the form  $C \cap A$  for some  $C \in \mathcal{C}$ . For a classical geometric example, if  
 124  $A$  is the set of four points on the Euclidean plane of the form  $(\pm 1, \pm 1)$ , then the  
 125 collection of all half-planes does not shatter  $A$ , the collection of all open balls does  
 126 not shatter  $A$ , but the collection of all convex sets shatters  $A$ . While  $A$  need not be  
 127 finite, it will usually be assumed to be so in Machine Learning applications. A finer  
 128 way to distinguish collections of sets that shatter a given set from those that do  
 129 not is by the *Vapnik-Chervonenkis dimension (VC-dimension)*, which is equal to  
 130 the cardinality of the largest finite set shattered by the collection, in case it exists,  
 131 or to infinity otherwise.

132 A concrete illustration of these ideas appears when considering threshold clas-  
 133 sifiers on the real line. Let  $\mathcal{H}$  be the collection of all indicator functions  $h_t$  given  
 134 by  $h_t(x) = 1$  if  $x \leq t$  and  $h_t(x) = 0$  otherwise. Each  $h_t$  is a Baire class 1 func-  
 135 tion, and the family  $\mathcal{H}$  is relatively compact in  $B_1(\mathbb{R})$ . In model-theoretic terms,  
 136  $\mathcal{H}$  is NIP, since no configuration of points and thresholds can realize the full inde-  
 137 pendence pattern of a binary matrix. By contrast, the family of parity functions  
 138  $\{x \mapsto (-1)^{\langle w, x \rangle} : w \in \{0, 1\}^n\}$  on  $\{0, 1\}^n$  (here  $\langle w, x \rangle$  is the usual vector dot product)  
 139 has the independence property and fails relative compactness in  $B_1(X)$ , capturing  
 140 the analytical meaning of instability. This dichotomy mirrors the behavior of con-  
 141 cept classes with finite versus infinite VC dimension in statistical learning theory.

Going back to the model theoretic framework, let

$$\mathcal{F}_\varphi(M) := \{\varphi(M, a) : a \in M^{|y|}\}$$

142 be the family of subsets of  $M^{|x|}$  defined by instances of the formula  $\varphi$ , where  
 143  $\varphi(M, a)$  is the set of  $|x|$ -tuples  $c$  in  $M$  for which  $M \models \varphi(c, a)$ . The fundamental  
 144 theorem of statistical learning states that a binary hypothesis class is PAC-learnable  
 145 if and only if it has finite VC-dimension, and the subsequent theorem connects the  
 146 rest of the concepts presented in this section.

147 **Theorem 2.1** (Laskowski). *The formula  $\varphi(x, y)$  has the NIP if and only if  $\mathcal{F}_\varphi(M)$   
 148 has finite VC-dimension.*

149 For two simple examples of formulas satisfying the NIP, consider first the lan-  
 150 guage  $L = \{<\}$  and the model  $M = (\mathbb{R}, <)$  of the reals with their usual linear order.  
 151 Take the formula  $\varphi(x, y)$  to mean  $x < y$ , then  $\varphi(M, a) = (-\infty, a)$ , and so  $\mathcal{F}_\varphi(M)$   
 152 is just the set of left open rays. The VC-dimension of this collection is 1, since it  
 153 can shatter a single point, but no two point set can be shattered since the rays are  
 154 downwards closed. Now in contrast, the collection of open intervals, given by the  
 155 formula  $\varphi(x; y_1, y_2) := (y_1 < x) \wedge (x < y_2)$ , has VC-dimension 2.

156 In this work, we study the corresponding notions of NIP (and hence PAC-  
 157 learnability) in the context of Compositional Computation Structures (CCS) in-  
 158 troduced in [ADIW24].

## 159        3. GENERAL TOPOLOGICAL PRELIMINARIES

160        In this section we give preliminaries from general topology and function space  
 161 theory. We include some of the proofs for completeness but a reader familiar with  
 162 these topics may skip them.

163        A *Polish space* is a separable and completely metrizable topological space. The  
 164 most important examples are the reals  $\mathbb{R}$ , the Cantor space  $2^{\mathbb{N}}$  (the set of all infinite  
 165 binary sequences, endowed with the product topology), and the Baire space  $\mathbb{N}^{\mathbb{N}}$  (the  
 166 set of all infinite sequences of naturals, also with the product topology). Countable  
 167 products of Polish spaces are Polish; this includes spaces like  $\mathbb{R}^{\mathbb{N}}$ , the space of  
 168 sequences of real numbers. A subspace of a Polish space is itself Polish if and only  
 169 if it is a  $G_{\delta}$ -set, that is, it can be written as the intersection of a countable family  
 170 of open subsets; in particular, closed subsets and open subsets of Polish spaces are  
 171 also Polish spaces.

172        In this work we talk a lot about subspaces, and so there is a pertinent subtlety  
 173 of the definitions worth mentioning: *completely metrizable space* is not the same  
 174 as *complete metric space*; for an illustrative example, notice that  $(0, 1)$  is home-  
 175 omorphic to the real line, and thus a Polish space (being Polish is a topological  
 176 property), but with the metric inherited from the reals, as a subspace,  $(0, 1)$  is **not**  
 177 a complete metric space. In summary, a Polish space has its topology generated by  
 178 *some* complete metric, but other metrics generating the same topology might not  
 179 be. In practice, such as when studying descriptive set theory, one finds that we can  
 180 often keep the metric implicit.

181        Given two topological spaces  $X$  and  $Y$  we denote by  $B_1(X, Y)$  the set of all func-  
 182 tions  $f : X \rightarrow Y$  such that for all open  $U \subseteq Y$ ,  $f^{-1}[U]$  is an  $F_{\sigma}$  subset of  $X$  (that  
 183 is, a countable union of closed sets); we call these types of functions *Baire class*  
 184 *1 functions*. When  $Y = \mathbb{R}$  we simply denote this collection by  $B_1(X)$ . We endow  
 185  $B_1(X, Y)$  with the topology of pointwise convergence (the topology inherited  
 186 from the product topology of  $Y^X$ ). By  $C_p(X, Y)$  we denote the set of all contin-  
 187 uous functions  $f : X \rightarrow Y$  with the topology of pointwise convergence. Similarly,  
 188  $C_p(X) := C_p(X, \mathbb{R})$ . A natural question is, how do topological properties of  $X$   
 189 translate to  $C_p(X)$  and vice versa? These questions, and in general the study of  
 190 these spaces, are the concern of  $C_p$ -theory, an active field of research in general  
 191 topology which was pioneered by A. V. Arhangel'skii and his students in the 1970's  
 192 and 1980's. This field has found many exciting applications in model theory and  
 193 functional analysis. Good recent surveys on the topics include [HT23] and [Tka11].  
 194 We begin with the following:

195        **Fact 3.1.** *If all open subsets of  $X$  are  $F_{\sigma}$  (in particular if  $X$  is metrizable), then*  
 196  $C_p(X, Y) \subseteq B_1(X, Y)$ .

197        The proof of the following fact (due to Baire) can be found in Section 10 of  
 198 [Tod97].

199        **Fact 3.2 (Baire).** *If  $X$  is a complete metric space, then the following are equivalent:*

- 200        (i)  *$f$  is a Baire class 1 function, that is,  $f \in B_1(X)$ .*
- 201        (ii)  *$f$  is a pointwise limit of continuous functions.*
- 202        (iii) *For every closed  $F \subseteq X$ , the restriction  $f|_F$  has a point of continuity.*

203        Moreover, if  $X$  is Polish and  $f \notin B_1(X)$ , then there exists countable  $D_0, D_1 \subseteq X$  and  
 204 reals  $a < b$  such that  $\overline{D_0} = \overline{D_1}$ ,  $D_0 \subseteq f^{-1}(-\infty, a]$  and  $D_1 \subseteq f^{-1}[b, \infty)$ .

205 A subset  $L \subseteq X$  is *relatively compact* in  $X$  if the closure of  $L$  in  $X$  is compact.  
 206 Relatively compact subsets of  $B_1(X)$  (for  $X$  Polish space) have been objects of  
 207 interest to many people working in Analysis and Topological Dynamics. We begin  
 208 with the following well-known result. Recall that a set  $A \subseteq \mathbb{R}^X$  of real-valued  
 209 functions is *pointwise bounded* if for every  $x \in X$  there is  $M_x > 0$  such that  $|f(x)| <$   
 210  $M_x$  for all  $f \in A$ . We include the proof for the reader's convenience:

211 **Lemma 3.3.** *Let  $X$  be a Polish space and  $A \subseteq B_1(X)$  be pointwise bounded. The  
 212 following are equivalent:*

- 213 (i)  $A$  is relatively compact in  $B_1(X)$ .
- 214 (ii)  $A$  is relatively countably compact in  $B_1(X)$ , i.e., every countable subset of  
 215  $A$  has an accumulation point in  $B_1(X)$ .
- 216 (iii)  $\overline{A} \subseteq B_1(X)$ , where  $\overline{A}$  denotes the closure in  $\mathbb{R}^X$ .

217 *Proof.* By definition, being pointwise bounded means that there is, for each  $x \in X$ ,  
 218  $M_x > 0$  such that, for every  $f \in A$ ,  $|f(x)| \leq M_x$ .

219 (i) $\Rightarrow$ (ii) holds in general.

220 (ii) $\Rightarrow$ (iii) Assume that  $A$  is relatively countably compact in  $B_1(X)$  and that  
 221  $f \in \overline{A} \setminus B_1(X)$ . By Fact 3.2, there are countable  $D_0, D_1 \subseteq X$  with  $\overline{D_0} = \overline{D_1}$ , and  
 222  $a < b$  such that  $D_0 \subseteq f^{-1}(-\infty, a]$  and  $D_1 \subseteq f^{-1}[b, \infty)$ . We claim that there is a  
 223 sequence  $\{f_n\}_{n \in \mathbb{N}} \subseteq A$  such that for all  $x \in D_0 \cup D_1$ ,  $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ . Indeed,  
 224 use the countability to enumerate  $D_0 \cup D_1$  as  $\{x_n\}_{n \in \mathbb{N}}$ . Then find, for each positive  
 225  $n$ ,  $f_n \in A$  with  $|f_n(x_i) - f(x_i)| < \frac{1}{n}$  for all  $i \leq n$ . The claim follows.

226 By relative countable compactness of  $A$ , there is an accumulation point  $g \in$   
 227  $B_1(X)$  of  $\{f_n\}_{n \in \mathbb{N}}$ . It is straightforward to show that since  $f$  and  $g$  agree on  $D_0 \cup D_1$ ,  
 228  $g$  does not have a point of continuity on the closed set  $\overline{D_0} = \overline{D_1}$ , which contradicts  
 229 Fact 3.2.

230 (iii) $\Rightarrow$ (i) Suppose that  $\overline{A} \subseteq B_1(X)$ . Then  $\overline{A} \cap B_1(X) = \overline{A}$  is a closed subset of  
 231  $\prod_{x \in X} [-M_x, M_x]$ ; Tychonoff's theorem states that the product of compact spaces  
 232 is always compact, and since closed subsets of compact spaces are compact,  $\overline{A}$  must  
 233 be compact, as desired.  $\square$

234 **3.1. From Rosenthal's dichotomy to NIP.** The fundamental idea that  
 235 connects the rich theory here presented to real-valued computations is the concept  
 236 of an *approximation*. In the reals, points of closure from some subset can always  
 237 be approximated by points inside the set, via a convergent sequence. For more  
 238 complicated spaces, such as  $C_p(X)$ , this fails in a remarkably intriguing way. Let  
 239 us show an example that is actually the protagonist of a celebrated result. Con-  
 240 sider the Cantor space  $X = 2^\mathbb{N}$  and let  $p_n(x) = x(n)$  define a continuous mapping  
 241  $X \rightarrow \{0, 1\}$ . Then one can show (see Chapter 1.1 of [Tod97] for details) that, per-  
 242 haps surprisingly, the only continuous functions in the closure of  $\{p_n\}_{n \in \mathbb{N}}$  are the  
 243 functions  $p_n$  themselves; moreover, none of the subsequences of  $\{p_n\}_{n \in \mathbb{N}}$  converge.  
 244 In some sense, this example is the worst possible scenario for convergence. The  
 245 topological space obtained from this closure is well-known. Topologists refer to it  
 246 as the Stone-Čech compactification of the discrete space of natural numbers, or  $\beta\mathbb{N}$   
 247 for short, and it is an important object of study in general topology.

248 **Theorem 3.4** (Rosenthal's Dichotomy). *If  $X$  is Polish and  $\{f_n\} \subseteq C_p(X)$  is point-  
 249 wise bounded, then either  $\{f_n\}_{n \in \mathbb{N}}$  contains a convergent subsequence or a subse-  
 250 quence whose closure (in  $\mathbb{R}^X$ ) is homeomorphic to  $\beta\mathbb{N}$ .*

251 In other words, a pointwise bounded set of continuous functions will either contain  
 252 a subsequence that converges or a subsequence whose closure is essentially  
 253 the same as the example mentioned in the previous paragraphs (the worst possible  
 254 scenario). Note that in the preceding example, the functions are trivially pointwise  
 255 bounded in  $\mathbb{R}^X$  as the functions can only take values 0 and 1.

256 If we intend to generalize our results from  $C_p(X)$  to the bigger space  $B_1(X)$ , we  
 257 find a similar dichotomy. Either every point of closure of the set of functions will  
 258 be a Baire class 1 function, or there is a sequence inside the set that behaves in the  
 259 worst possible way (which in this context, is the IP!). The theorem is usually not  
 260 phrased as a dichotomy but rather as an equivalence (with the NIP instead):

261 **Theorem 3.5** (Bourgain-Fremlin-Talagrand, Theorem 4G in [BFT78]). *Let  $X$  be  
 262 a Polish space and  $A \subseteq C_p(X)$  be pointwise bounded. The following are equivalent:*

- 263 (i)  *$A$  is relatively compact in  $B_1(X)$ , i.e.,  $\bar{A} \subseteq B_1(X)$ .*  
 263 (ii) *For every  $\{f_n\}_{n \in \mathbb{N}} \subseteq A$  and for every  $a < b$  there is  $I \subseteq \mathbb{N}$  such that*

$$\bigcap_{n \in I} f_n^{-1}(-\infty, a] \cap \bigcap_{n \notin I} f_n^{-1}[b, \infty) = \emptyset.$$

264 Our goal now is to characterize relatively compact subsets of  $B_1(X, Y)$  when  
 265  $Y = \mathbb{R}^P$  with  $P$  countable. Given  $P \in \mathcal{P}$  we denote the *projection map* onto the  
 266  $P$ -coordinate by  $\pi_P : \mathbb{R}^P \rightarrow \mathbb{R}$ . From a high-level topological interpretation, the  
 267 subsequent lemma states that, in this context, the spaces  $\mathbb{R}$  and  $\mathbb{R}^P$  are really not  
 268 that different, and that if we understand the Baire class 1 functions of one space,  
 269 then we also understand the functions of both. In fact,  $\mathbb{R}$  and any other Polish  
 270 space is embeddable as a closed subspace of  $\mathbb{R}^P$ .

271 **Lemma 3.6.** *Let  $X$  be a Polish space and  $\mathcal{P}$  be a countable set. Then,  $f \in B_1(X, \mathbb{R}^P)$   
 272 if and only if  $\pi_P \circ f \in B_1(X)$  for all  $P \in \mathcal{P}$ .*

Proof. Only one implication needs a proof. Suppose that  $\pi_P \circ f \in B_1(X)$  for all  $P \in \mathcal{P}$ . Let  $V$  be a basic open subset of  $\mathbb{R}^P$ . That is, there exists a finite  $\mathcal{P}' \subseteq \mathcal{P}$  such that  $V = \bigcap_{P \in \mathcal{P}'} \pi_P^{-1}[U_P]$  where  $U_P$  is open in  $\mathbb{R}$ . Finally,

$$f^{-1}[V] = \bigcap_{P \in \mathcal{P}'} (\pi_P \circ f)^{-1}[U_P]$$

273 is an  $F_\sigma$  set. Since  $\mathcal{P}$  is countable,  $\mathbb{R}^P$  is second countable so every open set  $U$  in  
 274  $\mathbb{R}^P$  is a countable union of basic open sets. Hence,  $f^{-1}[U]$  is  $F_\sigma$ .  $\square$

275 Below we consider  $\mathcal{P}$  with the discrete topology. For each  $f : X \rightarrow \mathbb{R}^P$  denote  
 276  $\hat{f}(P, x) := \pi_P \circ f(x)$  for all  $(P, x) \in \mathcal{P} \times X$ . Similarly, for each  $g : \mathcal{P} \times X \rightarrow \mathbb{R}$  denote  
 277  $\check{g}(x)(P) := g(P, x)$ . Given  $A \subseteq (\mathbb{R}^P)^X$ , we denote  $\hat{A}$  as the set of all  $\hat{f}$  such that  
 278  $f \in A$ .

279 The map  $(\mathbb{R}^P)^X \rightarrow \mathbb{R}^{P \times X}$  given by  $f \mapsto \hat{f}$  is a homeomorphism and its inverse is  
 280 given by  $g \mapsto \check{g}$ .

281 **Lemma 3.7.** *Let  $X$  be a Polish space and  $\mathcal{P}$  be countable. Then,  $f \in B_1(X, \mathbb{R}^P)$  if  
 282 and only if  $\hat{f} \in B_1(\mathcal{P} \times X)$ .*

Proof. ( $\Rightarrow$ ) Given an open set of reals  $U$ , we have that for every  $P \in \mathcal{P}$ ,  $f^{-1}[\pi_P^{-1}[U]]$   
 is  $F_\sigma$  by Lemma 3.6. Given that  $\mathcal{P}$  is a discrete countable space, we observe that

$$\hat{f}^{-1}[U] = \bigcup_{P \in \mathcal{P}} (\{P\} \times f^{-1}[\pi_P^{-1}[U]])$$

is also an  $F_\sigma$  set. ( $\Leftarrow$ ) By lemma 3.6 it suffices to show that  $\pi_P \circ f \in B_1(X)$  for all  $P \in \mathcal{P}$ . Fix an open  $U \subseteq \mathbb{R}$ . Write  $\hat{f}^{-1}[U] = \bigcup_{n \in \mathbb{N}} F_n$  where  $F_n$  is closed in  $\mathcal{P} \times X$ . Then,

$$(\pi_P \circ f)^{-1}[U] = \bigcup_{n \in \mathbb{N}} \{x \in X : (P, x) \in F_n\}$$

which is  $F_\sigma$ .  $\square$

We now direct our attention to a notion of the NIP that is more general than the one from the introduction. It can be interpreted as a sort of continuous version of the one presented in the preceding section.

**Definition 3.8.** We say that  $A \subseteq \mathbb{R}^X$  has the *Non-Independence Property* (NIP) if and only if for every  $\{f_n\}_{n \in \mathbb{N}} \subseteq A$  and for every  $a < b$  there are finite disjoint sets  $E, F \subseteq \mathbb{N}$  such that

$$\bigcap_{n \in E} f_n^{-1}(-\infty, a] \cap \bigcap_{n \in F} f_n^{-1}[b, \infty) = \emptyset.$$

Note that if  $X$  is compact and  $A \subseteq C_p(X)$ , then  $A$  has the NIP if and only if for every  $\{f_n\}_{n \in \mathbb{N}} \subseteq A$  and for every  $a < b$  there is  $I \subseteq \mathbb{N}$  such that

$$\bigcap_{n \in I} f_n^{-1}(-\infty, a] \cap \bigcap_{n \notin I} f_n^{-1}[b, \infty) = \emptyset.$$

Given  $A \subseteq Y^X$  and  $K \subseteq X$  we write  $A|_K := \{f|_K : f \in A\}$ , i.e., the set of all restrictions of functions in  $A$  to  $K$ . The following Theorem is a slightly more general version of Theorem 3.5.

**Theorem 3.9.** Assume that  $\mathcal{P}$  is countable,  $X$  is a Polish space, and  $A \subseteq C_p(X, \mathbb{R}^\mathcal{P})$  is such that  $\pi_P \circ A$  is pointwise bounded for all  $P \in \mathcal{P}$ . The following are equivalent for every compact  $K \subseteq X$ :

- (1)  $\overline{A|_K} \subseteq B_1(K, \mathbb{R}^\mathcal{P})$ .
- (2)  $\pi_P \circ A|_K$  has the NIP for every  $P \in \mathcal{P}$ .

*Proof.* (1) $\Rightarrow$ (2). Let  $P \in \mathcal{P}$ . Fix  $\{f_n\}_{n \in \mathbb{N}} \subseteq A$  and  $a < b$ . By (1) we have that  $\overline{A|_K} \subseteq B_1(K, \mathbb{R}^\mathcal{P})$ . Applying the homeomorphism  $f \mapsto \hat{f}$  and using lemma 3.7 we get  $\overline{\hat{A}|_{\mathcal{P} \times K}} \subseteq B_1(\mathcal{P} \times K)$ . By Theorem 3.5, there is  $I \subseteq \mathbb{N}$  such that

$$(\mathcal{P} \times K) \cap \bigcap_{n \in I} \hat{f}_n^{-1}(-\infty, a] \cap \bigcap_{n \notin I} \hat{f}_n^{-1}[b, \infty) = \emptyset$$

Hence,

$$K \cap \bigcap_{n \in I} (\pi_P \circ f_n)^{-1}(-\infty, a] \cap \bigcap_{n \notin I} (\pi_P \circ f_n)^{-1}[b, \infty) = \emptyset$$

By compactness, there are finite  $E \subseteq I$  and  $F \subseteq \mathbb{N} \setminus I$  such that

$$K \cap \bigcap_{n \in E} (\pi_P \circ f_n)^{-1}(-\infty, a] \cap \bigcap_{n \in F} (\pi_P \circ f_n)^{-1}[b, \infty) = \emptyset$$

Thus,  $\pi_P \circ A|_L$  has the NIP.

(2) $\Rightarrow$ (1) Fix  $f \in \overline{A|_K}$ . By lemma 3.6 it suffices to show that  $\pi_P \circ f \in B_1(K)$  for all  $P \in \mathcal{P}$ . By (2),  $\pi_P \circ A|_K$  has the NIP. Hence, by Theorem 3.5 we have  $\pi_P \circ A|_K \subseteq B_1(K)$ . But then  $\pi_P \circ f \in \overline{\pi_P \circ A|_K} \subseteq B_1(K)$ .  $\square$

299     Lastly, a simple but significant result that helps understand the operation of  
 300     restricting a set of functions to a specific subspace of the domain space  $X$ , of course  
 301     in the context of the NIP, is that we may always assume that said subspace is  
 302     closed. Concretely, whether we take its closure or not has no effect on the NIP:

303     **Lemma 3.10.** *Assume that  $X$  is Hausdorff and that  $A \subseteq C_p(X)$ . The following  
 304     are equivalent for every  $L \subseteq X$ :*

- 305       (i)  $A_L$  has the NIP.  
 306       (ii)  $A|_{\bar{L}}$  has the NIP.

*Proof.* It suffices to show that (i) $\Rightarrow$ (ii). Suppose that (ii) does not hold, i.e., that there are  $\{f_n\}_{n \in \mathbb{N}} \subseteq A$  and  $a < b$  such that for all finite disjoint  $E, F \subseteq \mathbb{N}$ :

$$\bar{L} \cap \bigcap_{n \in E} f_n^{-1}(-\infty, a] \cap \bigcap_{n \in F} f_n^{-1}[b, \infty) \neq \emptyset.$$

Pick  $a' < b'$  such that  $a < a' < b' < b$ . Then, for any finite disjoint  $E, F \subseteq \mathbb{N}$  we can choose

$$x \in \bar{L} \cap \bigcap_{n \in E} f_n^{-1}(-\infty, a') \cap \bigcap_{n \in F} f_n^{-1}(b', \infty)$$

By definition of closure:

$$L \cap \bigcap_{n \in E} f_n^{-1}(-\infty, a') \cap \bigcap_{n \in F} f_n^{-1}(b', \infty) \neq \emptyset.$$

307     This contradicts (i). □

#### 308     4. NIP IN THE CONTEXT OF COMPOSITIONAL COMPUTATION STRUCTURES

309     In this section, we study what the NIP tell us in the context of deep computations as defined in [ADIW24]. We say a structure  $(L, \mathcal{P}, \Gamma)$  is a *Compositional  
 310     Computation Structure* (CCS) if  $L \subseteq \mathbb{R}^{\mathcal{P}}$  is a subspace of  $\mathbb{R}^{\mathcal{P}}$ , with the pointwise  
 311     convergence topology, and  $\Gamma \subseteq L^L$  is a semigroup under composition. The motivation  
 312     for CCS comes from (continuous) model theory, where  $\mathcal{P}$  is a fixed collection  
 313     of predicates and  $L$  is a (real-valued) structure. Every point in  $L$  is identified with  
 314     its “type”, which is the tuple of all values the point takes on the predicates from  
 315      $\mathcal{P}$ , i.e., an element of  $\mathbb{R}^{\mathcal{P}}$ . In this context, elements of  $\mathcal{P}$  are called *features*. In the  
 316     discrete model theory framework, one views the space of complete-types as a sort of  
 317     compactification of the structure  $L$ . In this context, we don’t want to consider only  
 318     points in  $L$  (realized types) but in its closure  $\bar{L}$  (possibly unrealized types). The  
 319     problem is that the closure  $\bar{L}$  is not necessarily compact, an assumption that turns  
 320     out to be very useful in the context of continuous model theory. To bypass this  
 321     problem in a framework for deep computations, Alva, Dueñez, Iovino and Walton  
 322     introduced in [ADIW24] the concept of *shards*, which essentially consists in cover-  
 323     ing (a large fragment) of the space  $\bar{L}$  by compact, and hence pointwise-bounded,  
 324     subspaces (shards). We shall give the formal definition next.

326     A *sizer* is a tuple  $r_{\bullet} = (r_P)_{P \in \mathcal{P}}$  of positive real numbers indexed by  $\mathcal{P}$ . Given a  
 327     sizer  $r_{\bullet}$ , we define the  $r_{\bullet}$ -shard as:

$$L[r_{\bullet}] = L \cap \prod_{P \in \mathcal{P}} [-r_P, r_P]$$

328     For an illustrative example, we can frame Newton’s polynomial root approxima-  
 329     tion method in the context of a CCS (see Example 5.6 of [ADIW24] for details) as

follows. Begin by considering the extended complex numbers  $\hat{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$  with the usual Riemann sphere topology that makes it into a compact space (where unbounded sequences converge to  $\infty$ ). In fact, not only is this space compact but it is covered by the shard given by the sizer  $(1, 1, 1)$  (the unit sphere is contained in the cube  $[-1, 1]^3$ ). The space  $\hat{\mathbb{C}}$  is homeomorphic to the usual unit sphere  $S^2 := \{(x, y, z) : x^2 + y^2 + z^2 = 1\}$  of  $\mathbb{R}^3$ , by means of the stereographic projection and its inverse  $\hat{\mathbb{C}} \rightarrow S^2$ . This function is regarded as a triple of predicates  $x, y, z : \hat{\mathbb{C}} \rightarrow [-1, 1]$  where each will map an extended complex number to its corresponding real coordinate on the cube  $[-1, 1]^3$ . Now fix the cubic complex polynomial  $p(s) := s^3 - 1$ , and consider the map which performs one step in Newton's method at a particular (extended) complex number  $s$ , for finding a root of  $p$ ,  $\gamma_p : \hat{\mathbb{C}} \rightarrow \hat{\mathbb{C}}$ . The explicit inner workings of  $\gamma_p$  are irrelevant for this example, except for the fact that it is a continuous mapping. It follows that  $(S^2, \{x, y, z\}, \{\gamma_p^k : k \in \mathbb{N}\})$  is a CCS. The idea is that repeated applications of  $\gamma_p(s), \gamma_p \circ \gamma_p(s), \gamma_p \circ \gamma_p \circ \gamma_p(s), \dots$  would approximate a root of  $p$  provided  $s$  was a good enough initial guess.

The  $r_\bullet$ -type-shard is defined as  $\mathcal{L}[r_\bullet] = \overline{L[r_\bullet]}$  and  $\mathcal{L}_{sh}$  is the union of all type-shards. Notice that  $\mathcal{L}_{sh}$  is not necessarily equal to  $\mathcal{L} = \overline{L}$ , unless  $\mathcal{P}$  is countable (see [ADIW24]). A *transition* is a map  $f : L \rightarrow L$ , in particular, every element in the semigroup  $\Gamma$  is a transition (these are called *realized computations*). In practice, one would like to work with “definable” computations, i.e., ones that can be described by a computer. In this topological framework, being continuous is an expected requirement. However, as in the case of complete-types in model theory, we will work with “unrealized computations”, i.e., maps  $f : \mathcal{L}_{sh} \rightarrow \mathcal{L}_{sh}$ . Note that continuity of a computation does not imply that it can be continuously extended to  $\mathcal{L}_{sh}$ .

Suppose that a transition map  $f : L \rightarrow \mathcal{L}$  can be extended continuously to a map  $\mathcal{L} \rightarrow \mathcal{L}$ . Then, the Stone-Weierstrass theorem implies that the feature  $\pi_P \circ f$  (here  $P$  is a fixed predicate, and the feature is hence continuous) can be uniformly approximated by polynomials on the compact set  $\mathcal{L}[r_\bullet]$ . Theorem 2.2 in [ADIW24] formalizes the converse of this fact, in the sense that transitions maps that are not continuously extendable in this fashion cannot be obtained from simple constructions involving predicates. Under this framework, the features  $\pi_P \circ f$  of such transitions  $f$  are not approximable by polynomials, and so they are understood as “non-computable” since, again, we expect the operations computers carry out to be determined by elementary algebra corresponding to polynomials (namely addition and multiplication). Therefore it is crucial we assume some extendibility conditions.

We say that the CCS  $(L, \mathcal{P}, \Gamma)$  satisfies the *Extendibility Axiom* if for all  $\gamma \in \Gamma$ , there is  $\tilde{\gamma} : \mathcal{L}_{sh} \rightarrow \mathcal{L}_{sh}$  such that for every sizer  $r_\bullet$  there is an  $s_\bullet$  such that  $\tilde{\gamma}|_{\mathcal{L}[r_\bullet]} : \mathcal{L}[r_\bullet] \rightarrow \mathcal{L}[s_\bullet]$  is continuous. For a deeper discussion about this axiom, we refer the reader to [ADIW24].

A collection  $R$  of sizers is called *exhaustive* if  $\mathcal{L}_{sh} = \bigcup_{r_\bullet \in R} \mathcal{L}[r_\bullet]$ . We say that  $\Delta \subseteq \Gamma$  is  $R$ -*confined* if  $\gamma|_{\mathcal{L}[r_\bullet]} : \mathcal{L}[r_\bullet] \rightarrow \mathcal{L}[r_\bullet]$  for every  $r_\bullet \in R$  and  $\gamma \in \Delta$ . Elements in  $\Delta$  are called *real-valued computations* (in this article we will refer to them simply as *computations*) and elements in  $\bar{\Delta} \subseteq \mathcal{L}_{sh}^L$  are called (real-valued) *deep computations* or *ultracomputations*. By  $\tilde{\Delta}$  we denote the set of all extensions  $\tilde{\gamma}$  for  $\gamma \in \Delta$ . For a more complete description of this framework, we refer the reader to [ADIW24].

377 **4.1. NIP and Baire-1 definability of deep computations.** Under what con-  
 378 ditions are deep computations Baire class 1, and thus well-behaved according to  
 379 our framework, on type-shards? The next Theorem says that, again under the  
 380 assumption that  $\mathcal{P}$  is countable, the space of deep computations is a Rosenthal  
 381 compactum (when restricted to shards) if and only if the set of computations has  
 382 the NIP on features. Hence, we can import the theory of Rosenthal compacta into  
 383 this framework of deep computations.

384 **Theorem 4.1.** *Let  $(L, \mathcal{P}, \Gamma)$  be a CCS satisfying the Extendibility Axiom with  $\mathcal{P}$   
 385 countable. Let  $R$  be an exhaustive collection of sizers. Let  $\Delta \subseteq \Gamma$  be  $R$ -confined. The  
 386 following are equivalent.*

- 387 (1)  $\overline{\tilde{\Delta}}|_{\mathcal{L}[r_\bullet]} \subseteq B_1(\mathcal{L}[r_\bullet], \mathcal{L}[r_\bullet])$  for all  $r_\bullet \in R$ .  
 (2)  $\pi_P \circ \Delta|_{\mathcal{L}[r_\bullet]}$  has the NIP for all  $P \in \mathcal{P}$  and  $r_\bullet \in R$ , that is, for all  $P \in \mathcal{P}$ ,  
 $r_\bullet \in R$ ,  $a < b$ ,  $\{\gamma_n\}_{n \in \mathbb{N}} \subseteq \Delta$  there are finite disjoint  $E, F \subseteq \mathbb{N}$  such that

$$L[r_\bullet] \cap \bigcap_{n \in E} (\pi_P \circ \gamma_n)^{-1}(-\infty, a) \cap \bigcap_{n \in F} (\pi_P \circ \gamma_n)^{-1}[b, \infty) = \emptyset.$$

388 Moreover, if any (hence all) of the preceding conditions hold, then every deep  
 389 computation  $f \in \overline{\Delta}$  can be extended to a Baire-1 function on shards, i.e., there is  
 390  $\tilde{f} : \mathcal{L}_{sh} \rightarrow \mathcal{L}_{sh}$  such that  $\tilde{f}|_{\mathcal{L}[r_\bullet]} \in B_1(\mathcal{L}[r_\bullet], \mathcal{L}[r_\bullet])$  for all  $r_\bullet \in R$ . In particular, on  
 391 each shard every deep computation is the pointwise limit of a countable sequence of  
 392 computations.

393 *Proof.* Since  $\mathcal{P}$  is countable, then  $\mathcal{L}[r_\bullet] \subseteq \mathbb{R}^{\mathcal{P}}$  is Polish. Also, the Extendibility  
 394 Axiom implies that  $\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]}$  is a pointwise bounded set of continuous functions  
 395 for all  $P \in \mathcal{P}$ . Hence, Theorem 3.9 and Lemma 3.10 prove the equivalence of (1)  
 396 and (2). If (1) holds and  $f \in \overline{\Delta}$ , then write  $f = \underline{Ulim}_i \gamma_i$  as an ultralimit. Define  
 397  $\tilde{f} := \underline{Ulim}_i \tilde{\gamma}_i$ . Hence, for all  $r_\bullet \in R$  we have  $\tilde{f}|_{\mathcal{L}[r_\bullet]} \in \overline{\tilde{\Delta}}|_{\mathcal{L}[r_\bullet]} \subseteq B_1(\mathcal{L}[r_\bullet], \mathcal{L}[r_\bullet])$ . That  
 398 every deep computation is a pointwise limit of a countable sequence of computations  
 399 follows from the fact that for a Polish space  $X$  every compact subset of  $B_1(X)$  is  
 400 Fréchet-Urysohn (that is, a space where topological closures coincide with sequential  
 401 closures, see Theorem 3F in [BFT78] or Theorem 4.1 in [Deb13]).  $\square$

402 Given a countable set  $\Delta$  of computations satisfying the NIP on features and  
 403 shards (condition (2) of Theorem 4.1) we have that  $\overline{\tilde{\Delta}}|_{\mathcal{L}[r_\bullet]}$  (for a fixed sizer  $r_\bullet$ ) is  
 404 a separable *Rosenthal compactum* (compact subset of  $B_1(P \times \mathcal{L}[r_\bullet])$ ). The work of  
 405 Todorčević ([Tod99]) and Argyros, Dodos, Kanellopoulos ([ADK08]) culminates in  
 406 a trichotomy theorem for separable Rosenthal Compacta. Inspired by the work of  
 407 Glasner and Megrelishvili ([GM22]), we are interested to see how this allows us to  
 408 classify and obtain different levels of PAC-learnability (NIP).

409 Recall that a topological space  $X$  is *hereditarily separable* (HS) if every subspace  
 410 is separable and that  $X$  is *first countable* if every point in  $X$  has a countable local  
 411 basis. Every separable metrizable space is hereditarily separable and it is a result  
 412 of R. Pol that every hereditarily separable Rosenthal compactum is first countable  
 413 (see section 10 in [Deb13]). This suggests the following definition:

414 **Definition 4.2.** Let  $(L, \mathcal{P}, \Gamma)$  be a CCS satisfying the Extendibility Axiom and  $R$   
 415 be an exhaustive collection of sizers. Let  $\Delta \subseteq \Gamma$  be an  $R$ -confined countable set of  
 416 computations satisfying the NIP on shards and features (condition (2) in Theorem  
 417 4.1). We say that  $\Delta$  is:

- 418        (i) NIP<sub>1</sub> if  $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}}$  is first countable for every  $r_\bullet \in R$ .  
 419        (ii) NIP<sub>2</sub> if  $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}}$  is hereditarily separable for every  $r_\bullet \in R$ .  
 420        (iii) NIP<sub>3</sub> if  $\overline{\tilde{\Delta}|_{\mathcal{L}[r_\bullet]}}$  is metrizable for every  $r_\bullet \in R$ .

421        Observe that  $NIP_3 \Rightarrow NIP_2 \Rightarrow NIP_1 \Rightarrow NIP$ . A natural question that would  
 422        continue this work is to find examples of CCS that separate these levels of NIP. In  
 423        [Tod99], Todorčević isolates 3 canonical examples of Rosenthal compacta that wit-  
 424        ness the failure of the converse implications above. We now present some separable  
 425        and non-separable examples of Rosenthal compacta:

- 426        (1) *Alexandroff compactification of a discrete space of size continuum.* For each  
 427         $a \in 2^{\mathbb{N}}$  consider the map  $\delta_a : 2^{\mathbb{N}} \rightarrow \mathbb{R}$  given by  $\delta_a(x) = 1$  if  $x = a$  and  
 428         $\delta_a(x) = 0$  otherwise. Let  $A(2^{\mathbb{N}}) = \{\delta_a : a \in 2^{\mathbb{N}}\} \cup \{0\}$ , where 0 is the zero  
 429        map. Notice that  $A(2^{\mathbb{N}})$  is a compact subset of  $B_1(2^{\mathbb{N}})$ , in fact  $\{\delta_a : a \in 2^{\mathbb{N}}\}$   
 430        is a discrete subspace of  $B_1(2^{\mathbb{N}})$  and its pointwise closure is precisely  $A(2^{\mathbb{N}})$ .  
 431        Hence, this is a Rosenthal compactum which is not first countable. Notice  
 432        that this space is also not separable.  
 433        (2) *Extended Alexandroff compactification.* For each finite binary sequence  $s \in$   
 434         $2^{<\mathbb{N}}$ , let  $v_s : 2^{\mathbb{N}} \rightarrow \mathbb{R}$  be given by  $v_s(x) = 1$  if  $x$  extends  $s$  and  $v_s(x) =$   
 435        0 otherwise. Let  $\hat{A}(2^{\mathbb{N}})$  be the pointwise closure of  $\{v_s : s \in 2^{<\mathbb{N}}\}$ , i.e.,  
 436         $\hat{A}(2^{\mathbb{N}}) = A(2^{\mathbb{N}}) \cup \{v_s : s \in 2^{<\mathbb{N}}\}$ . Note that this space is a separable  
 437        Rosenthal compactum which is not first countable.  
 438        (3) *Split Cantor.* Let  $<$  be the lexicographic order in the space of infinite  
 439        binary sequences, i.e.,  $2^{\mathbb{N}}$ . For each  $a \in 2^{\mathbb{N}}$  let  $f_a^- : 2^{\mathbb{N}} \rightarrow \mathbb{R}$  be given by  
 440         $f_a^-(x) = 1$  if  $x < a$  and  $f_a^-(x) = 0$  otherwise. Let  $f_a^+ : 2^{\mathbb{N}} \rightarrow \mathbb{R}$  be given  
 441        by  $f_a^+(x) = 1$  if  $x \leq a$  and  $f_a^+(x) = 0$  otherwise. The split Cantor is the  
 442        space  $S(2^{\mathbb{N}}) = \{f_a^- : a \in 2^{\mathbb{N}}\} \cup \{f_a^+ : a \in 2^{\mathbb{N}}\}$ . This is a separable Rosenthal  
 443        compactum. One example of a countable dense subset is the set of all  $f_a^+$   
 444        and  $f_a^-$  where  $a$  is an infinite binary sequence that is eventually constant.  
 445        Moreover, it is hereditarily separable but it is not metrizable.  
 446        (4) *Alexandroff Duplicate.* Let  $K$  be any compact metric space and consider  
 447        the Polish space  $X = C(K) \sqcup K$ , i.e., the disjoint union of  $C(K)$  (with its  
 448        supremum norm topology) and  $K$ . For each  $a \in K$  define  $g_a^0, g_a^1 : X \rightarrow \mathbb{R}$  as  
 follows:

$$g_a^0(x) = \begin{cases} x(a), & x \in C(K) \\ 0, & x \in K \end{cases}$$

$$g_a^1(x) = \begin{cases} x(a), & x \in C(K) \\ \delta_a(x), & x \in K \end{cases}$$

446        Let  $D(K) = \{g_a^0 : a \in K\} \cup \{g_a^1 : a \in K\}$ . Notice that  $D(K)$  is a first  
 447        countable Rosenthal compactum. It is not separable if  $K$  is uncountable.  
 448        The interesting case will be when  $K = 2^{\mathbb{N}}$ .

- 449        (5) *Extended Alexandroff Duplicate of the split Cantor.* For each finite binary  
 450        sequence  $t \in 2^{<\mathbb{N}}$  let  $a_t \in 2^{\mathbb{N}}$  be the sequence starting with  $t$  and ending  
 451        with 0's and let  $b_t \in 2^{\mathbb{N}}$  be the sequence starting with  $t$  and ending with  
 452        1's. Define  $h_t : 2^{\mathbb{N}} \rightarrow \mathbb{R}$  by

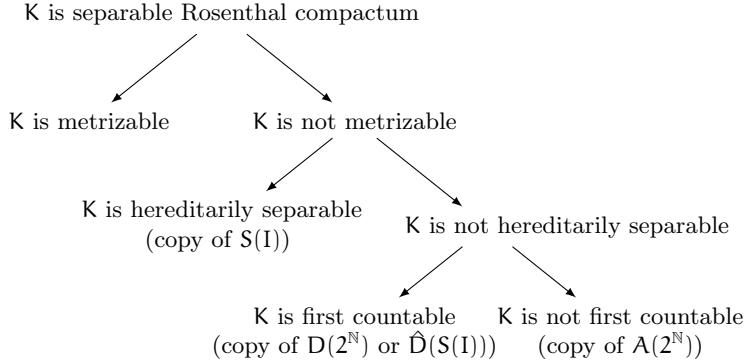
$$h_t(x) = \begin{cases} 0, & x < a_t \\ 1/2, & a_t \leq x \leq b_t \\ 1, & b_t < x \end{cases}$$

449 Let  $\hat{D}(S(2^{\mathbb{N}}))$  be the pointwise closure of the set  $\{h_t : t \in 2^{<\mathbb{N}}\}$ . Hence,  
 450  $\hat{D}(S(2^{\mathbb{N}}))$  is a separable first countable Rosenthal compactum which is not  
 451 hereditarily separable. In fact, it contains an uncountable discrete subspace  
 452 (see Theorem 5 in [Tod99]).

453 **Theorem 4.3** (Todorčević's Trichotomy, [Tod99], Theorem 3 in [ADK08]). *Let  $K$   
 454 be a separable Rosenthal Compactum.*

- 455 (i) *If  $K$  is hereditarily separable but non-metrizable, then  $S(2^{\mathbb{N}})$  embeds into  $K$ .*
- 456 (ii) *If  $K$  is first countable but not hereditarily separable, then either  $D(2^{\mathbb{N}})$  or  
 457  $\hat{D}(S(2^{\mathbb{N}}))$  embeds into  $K$ .*
- 458 (iii) *If  $K$  is not first countable, then  $A(2^{\mathbb{N}})$  embeds into  $K$ .*

459 In other words, we have the following classification:



460 461 Lastly, the definitions provided here for  $NIP_i$  ( $i = 1, 2, 3$ ) are topological.

462 **Question 4.4.** Is there a non-topological characterization for  $NIP_i$ ,  $i = 1, 2, 3$ ?

463 More can be said about the nature of the embeddings in Todorčević's Trichotomy.  
 464 Given a separable Rosenthal compactum  $K$ , there is typically more than one countable dense subset of  $K$ . We can view a separable Rosenthal compactum as the accumulation points of a countable family of pointwise bounded real-valued functions.  
 465 The choice of the countable families is not important when a bijection between  
 466 them can be lifted to a homeomorphism of their closures. To be more precise:  
 467

468 469 **Definition 4.5.** Given a Polish space  $X$ , a countable set  $I$  and two pointwise  
 470 bounded families  $\{f_i : i \in I\} \subseteq \mathbb{R}^X$ ,  $\{g_i : i \in I\} \subseteq \mathbb{R}^X$  indexed by  $I$ . We say that  
 471  $\{f_i : i \in I\}$  and  $\{g_i : i \in I\}$  are *equivalent* if and only if the map  $f_i \mapsto g_i$  is extended  
 472 to a homeomorphism from  $\overline{\{f_i : i \in I\}}$  to  $\overline{\{g_i : i \in I\}}$ .

473 Notice that in the separable examples discussed before ( $\hat{A}(2^{\mathbb{N}})$ ,  $S(2^{\mathbb{N}})$  and  $\hat{D}(S(2^{\mathbb{N}}))$ )  
 474 the countable dense subsets are indexed by the binary tree  $2^{<\mathbb{N}}$ . This choice of index  
 475 is useful because the Ramsey theory of perfect subsets of the Cantor space  $2^{\mathbb{N}}$   
 476 can be imported to analyze the behavior of the accumulation points. Since  $2^{<\mathbb{N}}$  is  
 477 countable, we can always choose this index for the countable dense subsets. This  
 478 is done in [ADK08].

479 **Definition 4.6.** Given a Polish space  $X$  and a pointwise bounded family  $\{f_t : t \in 2^{<\mathbb{N}}\}$ .  
 480 We say that  $\{f_t : t \in 2^{<\mathbb{N}}\}$  is *minimal* if and only if for every dyadic subtree  
 481  $\{s_t : t \in 2^{<\mathbb{N}}\}$  of  $2^{<\mathbb{N}}$ ,  $\{f_{s_t} : t \in 2^{<\mathbb{N}}\}$  is equivalent to  $\{f_t : t \in 2^{<\mathbb{N}}\}$ .

One of the main results in [ADK08] is that there are (up to equivalence) seven minimal families of Rosenthal compacta and that for every relatively compact  $\{f_t : t \in 2^{<\mathbb{N}}\} \subseteq B_1(X)$  there is a dyadic subtree  $\{s_t : t \in 2^{<\mathbb{N}}\}$  such that  $\{f_{s_t} : t \in 2^{<\mathbb{N}}\}$  is equivalent to one of the minimal families. We shall describe the minimal families next. We will follow the same notation as in [ADK08]. For any node  $t \in 2^{<\mathbb{N}}$ , we denote by  $t^\frown 0^\infty$  ( $t^\frown 1^\infty$ ) the infinite binary sequence starting with  $t$  and ending in 0's (1's). Fix a regular dyadic subtree  $R = \{s_t : t \in 2^{<\mathbb{N}}\}$  of  $2^{<\mathbb{N}}$  (i.e., a dyadic subtree such that every level of  $R$  is contained in a level of  $2^{<\mathbb{N}}$ ) with the property that for all  $s, s' \in R$ ,  $s^\frown 0^\infty \neq s'^\frown 0^\infty$  and  $s^\frown 1^\infty \neq s'^\frown 1^\infty$ . Given  $t \in 2^{<\mathbb{N}}$ , let  $v_t$  be the characteristic function of the set  $\{x \in 2^{\mathbb{N}} : x \text{ extends } t\}$ . Let  $<$  be the lexicographic order in  $2^{\mathbb{N}}$ . Given  $a \in 2^{\mathbb{N}}$ , let  $f_a^+ : 2^{\mathbb{N}} \rightarrow \{0, 1\}$  be the characteristic function of  $\{x \in 2^{\mathbb{N}} : a \leq x\}$  and let  $f_a^- : 2^{\mathbb{N}} \rightarrow \{0, 1\}$  be the characteristic function of  $\{x \in 2^{\mathbb{N}} : a < x\}$ . Given two maps  $f, g : 2^{\mathbb{N}} \rightarrow \mathbb{R}$  we denote by  $(f, g) : 2^{\mathbb{N}} \sqcup 2^{\mathbb{N}} \rightarrow \mathbb{R}$  the function which is  $f$  on the first copy of  $2^{\mathbb{N}}$  and  $g$  on the second copy of  $2^{\mathbb{N}}$ .

- (1)  $D_1 = \{\frac{1}{|t|+1}v_t : t \in 2^{<\mathbb{N}}\}$ . This is discrete in  $\overline{D_1} = A(2^{\mathbb{N}})$ .
- (2)  $D_2 = \{s_t^\frown 0^\infty : t \in 2^{<\mathbb{N}}\}$ . This is discrete in  $\overline{D_2} = 2^{\leq\mathbb{N}}$ .
- (3)  $D_3 = \{f_{s_t^\frown 0^\infty}^+ : t \in 2^{<\mathbb{N}}\}$ . This is a discrete in  $\overline{D_3} = S(2^{\mathbb{N}})$ .
- (4)  $D_4 = \{f_{s_t^\frown 1^\infty}^- : t \in 2^{<\mathbb{N}}\}$ . This is discrete in  $\overline{D_4} = S(2^{\mathbb{N}})$ .
- (5)  $D_5 = \{v_t : t \in 2^{<\mathbb{N}}\}$ . This is discrete in  $\overline{D_5} = \hat{A}(2^{\mathbb{N}})$ .
- (6)  $D_6 = \{(v_{s_t}, s_t^\frown 0^\infty) : t \in 2^{<\mathbb{N}}\}$ . This is discrete in  $\overline{D_6} = \hat{D}(2^{\mathbb{N}})$ .
- (7)  $D_7 = \{(v_{s_t}, f_{s_t^\frown 0^\infty}^+) : t \in 2^{<\mathbb{N}}\}$ . This is discrete in  $\overline{D_7} = \hat{D}(S(2^{\mathbb{N}}))$

**Theorem 4.7** (Heptacotomy of minimal families, Theorem 2 in [ADK08]). *Let  $X$  be Polish. For every relatively compact  $\{f_t : t \in 2^{<\mathbb{N}}\} \subseteq B_1(X)$ , there exists  $i = 1, 2, \dots, 7$  and a regular dyadic subtree  $\{s_t : t \in 2^{<\mathbb{N}}\}$  of  $2^{<\mathbb{N}}$  such that  $\{f_{s_t} : t \in 2^{<\mathbb{N}}\}$  is equivalent to  $D_i$ . Moreover, all  $D_i$  are minimal and mutually non-equivalent.*

**4.2. NIP and definability by universally measurable functions.** We now turn to the question: what happens when  $\mathcal{P}$  is uncountable? Notice that the countability assumption is crucial in the proof of Theorem 3.9 essentially because it makes  $\mathbb{R}^{\mathcal{P}}$  a Polish space. For the uncountable case, we may lose Baire-1 definability so we shall replace  $B_1(X)$  by a bigger class. Recall that the purpose of studying the class of Baire-1 functions is that a pointwise limit of continuous functions is not necessarily continuous. In [BFT78], J. Bourgain, D.H. Fremlin and M. Talagrand characterized the Non-Independence Property of a set of continuous functions with various notions of compactness in function spaces containing  $C(X)$ , such as  $B_1(X)$ . In this section we will replace  $B_1(X)$  with the larger space  $M_r(X)$  of universally measurable functions. The development of this section is based on Theorem 2F in [BFT78]. We now give the relevant definitions. Readers with little familiarity with measure theory can review the appendix for standard definitions appearing in this subsection.

Given a Hausdorff space  $X$  and a measurable space  $(Y, \Sigma)$ , we say that  $f : X \rightarrow Y$  is *universally measurable* (with respect to  $\Sigma$ ) if  $f^{-1}(E)$  is universally measurable for every  $E \in \Sigma$ , i.e.,  $f^{-1}(E)$  is  $\mu$ -measurable for every Radon probability measure  $\mu$  on  $X$ . When  $Y = \mathbb{R}$  we will always take  $\Sigma = \mathcal{B}(\mathbb{R})$ , the Borel  $\sigma$ -algebra of  $\mathbb{R}$ . In that case, a function  $f : X \rightarrow \mathbb{R}$  is universally measurable if and only if  $f^{-1}(U)$  is  $\mu$ -measurable for every Radon probability measure  $\mu$  on  $X$  and every open set  $U \subseteq \mathbb{R}$ . Following [BFT78], the collection of all universally measurable real-valued functions will be denoted by  $M_r(X)$ . In the context of deep computations, we will

be interested in transition maps from a state space  $L \subseteq \mathbb{R}^P$  to itself. There are two natural  $\sigma$ -algebras one can consider in the product space  $\mathbb{R}^P$ : the Borel  $\sigma$ -algebra, i.e., the  $\sigma$ -algebra generated by open sets in  $\mathbb{R}^P$ ; and the cylinder  $\sigma$ -algebra, i.e., the  $\sigma$ -algebra generated by Borel cylinder sets or equivalently basic open sets in  $\mathbb{R}^P$ . Note that when  $P$  is countable, both  $\sigma$ -algebras coincide but in general the cylinder  $\sigma$ -algebra is strictly smaller. We will use the cylinder  $\sigma$ -algebra to define universally measurable maps  $f : \mathbb{R}^P \rightarrow \mathbb{R}^P$ . The reason for this choice is because of the following characterization:

**Lemma 4.8.** *Let  $X$  be a Hausdorff space and  $Y = \prod_{i \in I} Y_i$  be any product of measurable spaces  $(Y_i, \Sigma_i)$  for  $i \in I$ . Let  $\Sigma_Y$  be the cylinder  $\sigma$ -algebra generated by the measurable spaces  $(Y_i, \Sigma_i)$ . Let  $f : X \rightarrow Y$ . The following are equivalent:*

- (i)  $f : X \rightarrow Y$  is universally measurable (with respect to  $\Sigma_Y$ ).
- (ii)  $\pi_i \circ f : X \rightarrow Y_i$  is universally measurable (with respect to  $\Sigma_i$ ) for all  $i \in I$ .

*Proof.* (i)  $\Rightarrow$  (ii) is clear since the projection maps  $\pi_i$  are measurable and the composition of measurable functions is measurable. To prove (ii)  $\Rightarrow$  (i), suppose that  $C = \prod_{i \in I} C_i$  is a measurable cylinder and let  $J$  be the finite set of  $i \in I$  such that  $C_i \neq Y_i$ . Then,  $C = \bigcap_{i \in J} \pi_i^{-1}(C_i)$  so  $f^{-1}(C) = \bigcap_{i \in J} (\pi_i \circ f)^{-1}(C_i)$  is a universally measurable set by assumption.  $\square$

The previous lemma says that a transition map is universally measurable if and only if it is universally measurable on all its features. In other words, we can check measurability of a transition just by checking measurability in all its features. We will denote by  $M_r(X, \mathbb{R}^P)$  the collection of all universally measurable functions  $f : X \rightarrow \mathbb{R}^P$  (with respect to the cylinder  $\sigma$ -algebra), endowed with the topology of pointwise convergence.

**Definition 4.9.** Let  $(L, P, \Gamma)$  be a CCS. We say that a transition  $f : L \rightarrow L$  is *universally measurable shard-definable* if and only if there exists  $\tilde{f} : \mathcal{L}_{sh} \rightarrow \mathcal{L}_{sh}$  extending  $f$  such that for every sizer  $r_\bullet$  there is a sizer  $s_\bullet$  such that the restriction  $\tilde{f}|_{\mathcal{L}[r_\bullet]} : \mathcal{L}[r_\bullet] \rightarrow \mathcal{L}[s_\bullet]$  is universally measurable, i.e.  $\pi_P \circ \tilde{f}|_{\mathcal{L}[r_\bullet]} : \mathcal{L}[r_\bullet] \rightarrow [-s_P, s_P]$  is  $\mu$ -measurable for every Radon probability measure  $\mu$  on  $\mathcal{L}[r_\bullet]$ .

We will need the following result about NIP and universally measurable functions:

**Theorem 4.10** (Bourgain-Fremlin-Talagrand, Theorem 2F in [BFT78]). *Let  $X$  be a Hausdorff space and  $A \subseteq C(X)$  be pointwise bounded. The following are equivalent:*

- (i)  $\overline{A} \subseteq M_r(X)$ .
- (ii) For every compact  $K \subseteq X$ ,  $A|_K$  has the NIP.
- (iii) For every Radon measure  $\mu$  on  $X$ ,  $A$  is relatively countably compact in  $\mathcal{L}^0(X, \mu)$ , i.e., every countable subset of  $A$  has an accumulation point in  $\mathcal{L}^0(X, \mu)$ .

Theorem 3.5 immediately yields the following.

**Theorem 4.11.** *Let  $(L, P, \Gamma)$  be a CCS satisfying the Extendibility Axiom. Let  $R$  be an exhaustive collection of sizers. Let  $\Delta \subseteq \Gamma$  be  $R$ -confined. If  $\pi_P \circ \Delta|_{L[r_\bullet]}$  has the NIP for all  $P \in P$  and all  $r_\bullet \in R$ , then every deep computation is universally measurable shard-definable.*

572 *Proof.* By the Extendibility Axiom, Theorem 3.5 and lemma 3.10 we have that  
 573  $\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]} \subseteq M_r(\mathcal{L}[r_\bullet])$  for all  $r_\bullet \in R$  and  $P \in \mathcal{P}$ . Let  $f \in \overline{\Delta}$  be a deep computation.  
 574 Write  $f = \mathcal{U} \lim_i \gamma_i$  as an ultralimit of computations in  $\Delta$ . Define  $\tilde{f} := \mathcal{U} \lim_i \tilde{\gamma}_i$ .  
 575 Then, for all  $r_\bullet \in R$  and  $P \in \mathcal{P}$   $\pi_P \circ \tilde{\gamma}_i|_{\mathcal{L}[r_\bullet]} \in M_r(\mathcal{L}[r_\bullet])$  for all  $i$  so  $\pi_P \circ f|_{\mathcal{L}[r_\bullet]} \in$   
 576  $\pi_P \circ \tilde{\Delta}|_{\mathcal{L}[r_\bullet]} \subseteq M_r(\mathcal{L}[r_\bullet])$ .  $\square$

577 **Question 4.12.** Under the same assumptions of the previous Theorem, suppose  
 578 that every deep computation of  $\Delta$  is universally measurable shard-definable. Must  
 579  $\pi_P \circ \Delta|_{L[r_\bullet]}$  have the NIP for all  $P \in \mathcal{P}$  and all  $r_\bullet \in R$ ?

580 **4.3. Talagrand stability and definability by universally measurable func-**  
 581 **tions.** There is another notion closely related to NIP, introduced by Talagrand  
 582 in [Tal84] while studying Pettis integration. Suppose that  $X$  is a compact Haus-  
 583 dorff space and  $A \subseteq \mathbb{R}^X$ . Let  $\mu$  be a Radon probability measure on  $X$ . Given a  
 584  $\mu$ -measurable set  $E \subseteq X$ , a positive integer  $k$  and real numbers  $a < b$ . we write:

$$D_k(A, E, a, b) = \bigcup_{f \in A} \{x \in E^{2k} : f(x_{2i}) \leq a, f(x_{2i+1}) \geq b \text{ for all } i < k\}$$

585 We say that  $A$  is *Talagrand  $\mu$ -stable* if and only if for every  $\mu$ -measurable  
 586 set  $E \subseteq X$  of positive measure and for every  $a < b$  there is  $k \geq 1$  such that  
 587  $(\mu^{2k})^*(D_k(A, E, a, b)) < (\mu(E))^{2k}$ . Notice that we work with the outer measure  
 588 because it is not necessarily true that the sets  $D_k(A, E, a, b)$  are  $\mu$ -measurable.  
 589 This is certainly the case when  $A$  is a countable set of continuous (or  $\mu$ -measurable)  
 590 functions.

591 The following lemma establishes that Talagrand stability is a way to ensure that  
 592 deep computations are definable by measurable functions. We include the proof for  
 593 the reader's convenience.

594 **Lemma 4.13.** *If  $A$  is Talagrand  $\mu$ -stable, then  $\overline{A}$  is also Talagrand  $\mu$ -stable and  
 595  $\overline{A} \subseteq \mathcal{L}^0(X, \mu)$ .*

596 *Proof.* First, observe that a subset of a  $\mu$ -stable set is  $\mu$ -stable. To show that  $\overline{A}$   
 597 is  $\mu$ -stable, observe that  $D_k(\overline{A}, E, a, b) \subseteq D_k(A, E, a', b')$  where  $a < a' < b' <$   
 598  $b$  and  $E$  is a  $\mu$ -measurable set with positive measure. It suffices to show that  
 599  $\overline{A} \subseteq \mathcal{L}^0(X, \mu)$ . Suppose that there exists  $f \in \overline{A}$  such that  $f \notin \mathcal{L}^0(X, \mu)$ . By a  
 600 characterization of measurable functions (see 413G in [Fre03]), there exists a  $\mu$ -  
 601 measurable set  $E$  of positive measure and  $a < b$  such that  $\mu^*(P) = \mu^*(Q) = \mu(E)$   
 602 where  $P = \{x \in E : f(x) \leq a\}$  and  $Q = \{x \in E : f(x) \geq b\}$ . Then, for any  $k \geq 1$ :  
 603  $(P \times Q)^k \subseteq D_k(\{f\}, E, a, b)$  so  $(\mu^{2k})^*(D_k(\{f\}, E, a, b)) = (\mu^*(P)\mu^*(Q))^k = (\mu(E))^{2k}$ .  
 604 Thus,  $\{f\}$  is not  $\mu$ -stable, but we argued before that a subset of a  $\mu$ -stable set must  
 605 be  $\mu$ -stable.  $\square$

606 We say that  $A$  is *universally Talagrand stable* if  $A$  is Talagrand  $\mu$ -stable for  
 607 every Radon probability measure  $\mu$  on  $X$ . A similar argument as before, yields the  
 608 following:

609 **Theorem 4.14.** *Let  $(L, \mathcal{P}, \Gamma)$  be a CCS satisfying the Extendibility Axiom. If  
 610  $\pi_P \circ \Delta|_{L[r_\bullet]}$  is universally Talagrand stable for all  $P \in \mathcal{P}$  and all sizers  $r_\bullet$ , then  
 611 every deep computation is universally measurable sh-definable.*

612 It is then natural to ask: what is the relationship between Talagrand stability  
 613 and the NIP? The following dichotomy will be useful.

614 **Lemma 4.15** (Fremlin's Dichotomy, 463K in [Fre03]). *If  $(X, \Sigma, \mu)$  is a perfect  $\sigma$ -  
 615 finite measure space (in particular, for  $X$  compact and  $\mu$  a Radon probability measure  
 616 on  $X$ ) and  $\{f_n : n \in \mathbb{N}\}$  be a sequence of real-valued measurable functions on  $X$ , then  
 617 either:*

- 618 (i)  $\{f_n : n \in \mathbb{N}\}$  has a subsequence that converges  $\mu$ -almost everywhere, or  
 619 (ii)  $\{f_n : n \in \mathbb{N}\}$  has a subsequence with no  $\mu$ -measurable accumulation point in  
 620  $\mathbb{R}^X$ .

621 The preceding lemma can be considered as the measure theoretic version of  
 622 Rosenthal's Dichotomy. Combining this dichotomy with the Theorem 4.10 we get  
 623 the following result:

624 **Theorem 4.16.** *Let  $X$  be a Hausdorff space and  $A \subseteq C(X)$  be pointwise bounded.  
 625 The following are equivalent:*

- 626 (i)  $\overline{A} \subseteq M_r(X)$ .  
 627 (ii) For every compact  $K \subseteq X$ ,  $A|_K$  has the NIP.  
 628 (iii) For every Radon measure  $\mu$  on  $X$ ,  $A$  is relatively countably compact in  
 629  $L^0(X, \mu)$ , i.e., every countable subset of  $A$  has an accumulation point in  
 630  $L^0(X, \mu)$ .  
 631 (iv) For every Radon measure  $\mu$  on  $X$  and every sequence  $\{f_n : n \in \mathbb{N}\} \subseteq A$ ,  
 632 there is a subsequence that converges  $\mu$ -almost everywhere.

633 *Proof.* Notice that the equivalence (i)-(iii) is Theorem 4.10. Notice that the equivalence  
 634 of (iii) and (iv) is Fremlin's Dichotomy Theorem.  $\square$

635 **Lemma 4.17.** *Let  $X$  be a compact Hausdorff space and  $A \subseteq C(X)$  be pointwise  
 636 bounded. If  $A$  is universally Talagrand stable, then  $A$  has the NIP.*

637 *Proof.* By Theorem 4.10, it suffices to show that  $A$  is relatively countably compact  
 638 in  $L^0(X, \mu)$  for all Radon probability measure  $\mu$  on  $X$ . Since  $A$  is Talagrand  $\mu$ -stable  
 639 for any such  $\mu$ , then  $\overline{A} \subseteq L^0(X, \mu)$ . In particular,  $A$  is relatively countably compact  
 640 in  $L^0(X, \mu)$ .  $\square$

641 **Question 4.18.** Is the converse true?

642 There is a delicate point in this question, as it may be sensitive to set-theoretic  
 643 axioms (even assuming countability of  $A$ ).

644 **Theorem 4.19** (Talagrand, Theorem 9-3-1(a) in [Tal84]). *Let  $X$  be a compact  
 645 Hausdorff space and  $A \subseteq M_r(X)$  be countable and pointwise bounded. Assume that  
 646  $[0, 1]$  is not the union of  $< c$  closed measure zero sets. If  $A$  has the NIP, then  $A$  is  
 647 universally Talagrand stable.*

648 **Theorem 4.20** (Fremlin, Shelah, [FS93]). *It is consistent that there exists a countable  
 649 pointwise bounded set of Lebesgue measurable functions with the NIP which is  
 650 not Talagrand stable with respect to Lebesgue measure.*

## 651 APPENDIX: MEASURE THEORY

652 Given a set  $X$ , a collection  $\Sigma$  of subsets of  $X$  is called a  $\sigma$ -algebra if  $\Sigma$  contains  
 653  $X$  and is closed under complements and countable unions. Hence, for example, a

654  $\sigma$ -algebra is also closed under countable intersections. Intuitively, a  $\sigma$ -algebra is  
 655 a collection of sets in which we can define a  $\sigma$ -additive measure. We call sets in  
 656 a  $\sigma$ -algebra  $\Sigma$  *measurable sets* and the pair  $(X, \Sigma)$  a measurable space. If  $X$  is a  
 657 topological space, there is a natural  $\sigma$ -algebra of subsets of  $X$ , namely the *Borel*  
 658  $\sigma$ -*algebra*  $\mathcal{B}(X)$ , i.e., the smallest  $\sigma$ -algebra containing all open subsets of  $X$ . Given  
 659 two measurable spaces  $(X, \Sigma_X)$  and  $(Y, \Sigma_Y)$ , we say that a function  $f : X \rightarrow Y$  is  
 660 *measurable* if and only if  $f^{-1}(E) \in \Sigma_X$  for every  $E \in \Sigma_Y$ . In particular, we say that  
 661  $f : X \rightarrow \mathbb{R}$  is measurable if  $f^{-1}(E) \in \Sigma_X$  for all  $E \in \mathcal{B}(\mathbb{R})$  (equivalently,  $E$  open in  
 662  $\mathbb{R}$ ).

663 Given a measurable space  $(X, \Sigma)$ , a  *$\sigma$ -additive measure* is a non-negative function  
 664  $\mu : \Sigma \rightarrow \mathbb{R}$  with the property that  $\mu(\emptyset) = 0$  and  $\mu(\bigcup_{n=0}^{\infty} A_n) = \sum_{n=0}^{\infty} \mu(A_n)$   
 665 whenever  $\{A_n : n \in \mathbb{N}\} \subseteq \Sigma$  is pairwise disjoint. We call  $(X, \Sigma, \mu)$  a *measure space*.  
 666 A  $\sigma$ -additive measure is called a *probability measure* if  $\mu(X) = 1$ . A measure  $\mu$   
 667 is *complete* if for every  $A \subseteq B \in \Sigma$ ,  $\mu(B) = 0$  implies  $A \in \Sigma$ . In words, subsets  
 668 of measure-zero sets are always measurable (and hence, by the monotonicity of  
 669  $\mu$ , have measure zero as well). A measure  $\mu$  is  *$\sigma$ -finite* if  $X = \bigcup_{n=1}^{\infty} X_n$  where  
 670  $\mu(X_n) < \infty$  for all  $n \in \mathbb{N}$  (i.e.,  $X$  can be decomposed into countably many finite  
 671 measure sets). A measure  $\mu$  is *perfect* if for every measurable  $f : X \rightarrow \mathbb{R}$  and  
 672 every measurable set  $E$  with  $\mu(E) > 0$ , there exists a compact  $K \subseteq f(E)$  such that  
 673  $\mu(f^{-1}(K)) > 0$ . We say that a property  $\phi(x)$  about  $x \in X$  holds  $\mu$ -*almost everywhere*  
 674 if  $\mu(\{x \in X : \phi(x) \text{ does not hold}\}) = 0$ .

675 A special example of the preceding concepts is that of a *Radon measure*. If  $X$  is  
 676 a Hausdorff topological space, then a measure  $\mu$  on the Borel sets of  $X$  is called a  
 677 *Radon measure* if

- 678 • for every open set  $U$ ,  $\mu(U)$  is the supremum of  $\mu(K)$  over all compact  $K \subseteq U$ ,  
 679 that is, the measure of open sets may be approximated via compact sets;  
 680 and
- 681 • every point of  $X$  has a neighborhood  $U \ni x$  for which  $\mu(U)$  is finite.

682 Perhaps the most famous example of a Radon measure on  $\mathbb{R}$  is the Lebesgue  
 683 measure of Borel sets. If  $X$  is finite,  $\mu(A) := |A|$  (the cardinality of  $A$ ) defines a  
 684 Radon measure on  $X$ . Every Radon measure is perfect (see 451A, 451B and 451C  
 685 in [Fre03]).

686 While not immediately obvious, sets can be measurable according to one mea-  
 687 sure, but non-measurable according to another. Given a measure space  $(X, \Sigma, \mu)$   
 688 we say that a set  $E \subseteq X$  is  $\mu$ -*measurable* if there are  $A, B \in \Sigma$  such that  $A \subseteq E \subseteq B$   
 689 and  $\mu(B \setminus A) = 0$ . The set of all  $\mu$ -measurable sets is a  $\sigma$ -algebra containing  $\Sigma$  and  
 690 it is denoted by  $\Sigma_{\mu}$ . A set  $E \subseteq X$  is *universally measurable* if it is  $\mu$ -measurable for  
 691 every Radon probability measure on  $X$ . It follows that Borel sets are universally  
 692 measurable. We say that  $f : X \rightarrow \mathbb{R}$  is  $\mu$ -*measurable* if  $f^{-1}(E) \in \Sigma_{\mu}$  for all  $E \in \mathcal{B}(\mathbb{R})$   
 693 (equivalently,  $E$  open in  $\mathbb{R}$ ). The set of all  $\mu$ -measurable functions is denoted by  
 694  $\mathcal{L}^0(X, \mu)$ .

Recall that if  $\{X_i : i \in I\}$  is a collection of topological spaces indexed by some  
 set  $I$ , then the product space  $X := \prod_{i \in I} X_i$  is endowed with the topology generated  
 by *cylinders*, that is, sets of the form  $\prod_{i \in I} U_i$  where each  $U_i$  is open in  $X_i$ , and  
 $U_i = X_i$  except for finitely many indices  $i \in I$ . If each space is measurable, say we  
pair  $X_i$  with a  $\sigma$ -algebra  $\Sigma_i$ , then there are multiple ways to interpret the product  
space  $X$  as a measurable space, but the interpretation we care about in this paper  
is the so called *cylinder  $\sigma$ -algebra*, as used in Lemma 4.8. Namely, let  $\Sigma$  be the

$\sigma$ -algebra generated by sets of the form

$$\prod_{i \in I} C_i, \quad C_i \in \Sigma_i, \quad C_i = X_i \text{ for all but finitely many } i \in I.$$

695     We remark that when  $I$  is uncountable and  $\Sigma_i = \mathcal{B}(X_i)$  for all  $i \in I$ , then  $\Sigma$  is,  
696     in general, strictly **smaller** than  $\mathcal{B}(X)$ .

697

## REFERENCES

- 698 [ADIW24] Samson Alva, Eduardo Dueñez, Jose Iovino, and Claire Walton. Approximability of  
699     deep equilibria. *arXiv preprint arXiv:2409.06064*, 2024. Last revised 20 May 2025,  
700     version 3.
- 701 [ADK08] Spiros A. Argyros, Pandelis Dodos, and Vassilis Kanellopoulos. A classification of sepa-  
702     rable Rosenthal compacta and its applications. *Dissertationes Mathematicae*, 449:1–55,  
703     2008.
- 704 [Ark91] A. V. Arkhangel'skii. *Topological Function Spaces*. Springer, New York, 1st edition,  
705     1991.
- 706 [BD19] Shai Ben-David. Understanding machine learning through the lens of model theory.  
707     *The Bulletin of Symbolic Logic*, 25(3):319–340, 2019.
- 708 [BFT78] J. Bourgain, D. H. Fremlin, and M. Talagrand. Pointwise compact sets of Baire-  
709     measurable functions. *American Journal of Mathematics*, 100(4):845–886, 1978.
- 710 [BKK] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. Deep equilibrium models. Preprint.  
711     <https://arxiv.org/abs/1909.01377>, <https://implicit-layers-tutorial.org/>.
- 712 [CRBD] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural  
713     ordinary differential equations. Preprint. <https://arxiv.org/abs/1806.07366>.
- 714 [Deb13] Gabriel Debs. Descriptive aspects of Rosenthal compacta. In *Recent Progress in Gen-  
715     eral Topology III*, pages 205–227. Springer, 2013.
- 716 [Fre00] David H. Fremlin. *Measure Theory, Volume 1: The Irreducible Minimum*. Torres  
717     Fremlin, Colchester, UK, 2000. Second edition 2011.
- 718 [Fre01] David H. Fremlin. *Measure Theory, Volume 2: Broad Foundations*. Torres Fremlin,  
719     Colchester, UK, 2001.
- 720 [Fre03] David H. Fremlin. *Measure Theory, Volume 4: Topological Measure Spaces*. Torres  
721     Fremlin, Colchester, UK, 2003.
- 722 [FS93] David H. Fremlin and Saharon Shelah. Pointwise compact and stable sets of measurable  
723     functions. *Journal of Symbolic Logic*, 58(2):435–455, 1993.
- 724 [GM22] Eli Glasner and Michael Megrelishvili. Todorčević' trichotomy and a hierarchy in the  
725     class of tame dynamical systems. *Transactions of the American Mathematical Society*,  
726     375(7):4513–4548, 2022.
- 727 [Gro52] A. Grothendieck. Critères de compacité dans les espaces fonctionnels généraux. *Amer.  
728     J. Math.*, 74:168–186, 1952.
- 729 [HT23] Clovis Hamel and Franklin D. Tall.  $C_p$ -theory for model theorists. In Jose Iovino, editor,  
730     *Beyond First Order Model Theory, Volume II*, chapter 5, pages 176–213. Chapman  
731     and Hall/CRC, 2023.
- 732 [Kec95] Alexander S. Kechris. *Classical Descriptive Set Theory*, volume 156 of *Graduate Texts  
733     in Mathematics*. Springer-Verlag, 1995.
- 734 [Kha20] Karim Khanaki. Stability, nip, and nsop; model theoretic properties of formulas via  
735     topological properties of function spaces. *Mathematical Logic Quarterly*, 66(2):136–  
736     149, 2020.
- 737 [She71] Saharon Shelah. Stability, the f.c.p., and superstability; model theoretic properties of  
738     formulas in first order theory. *Ann. Math. Logic*, 3(3):271–362, 1971.
- 739 [She78] Saharon Shelah. *Unstable Theories*, volume 187 of *Studies in Logic and the Foun-  
740     dations of Mathematics*. North-Holland, 1978.
- 741 [She90] Saharon Shelah. *Classification theory and the number of nonisomorphic models*, vol-  
742     ume 92 of *Studies in Logic and the Foundations of Mathematics*. North-Holland Pub-  
743     lishing Co., Amsterdam, second edition, 1990.
- 744 [Sim15a] Pierre Simon. *A Guide to NIP Theories*, volume 44 of *Lecture Notes in Logic*. Cam-  
745     bridge University Press, 2015.

- 746 [Sim15b] Pierre Simon. Rosenthal compacta and NIP formulas. *Fundamenta Mathematicae*,  
747 231(1):81–92, 2015.
- 748 [Tal84] Michel Talagrand. *Pettis Integral and Measure Theory*, volume 51 of *Memoirs of  
749 the American Mathematical Society*. American Mathematical Society, Providence, RI,  
750 USA, 1984. Includes bibliography (pp. 220–224) and index.
- 751 [Tal87] Michel Talagrand. The Glivenko-Cantelli problem. *The Annals of Probability*,  
752 15(3):837–870, 1987.
- 753 [Tka11] Vladimir V. Tkachuk. *A Cp-Theory Problem Book: Topological and Function Spaces*.  
754 Problem Books in Mathematics. Springer, 2011.
- 755 [Tod97] Stevo Todorcevic. *Topics in Topology*, volume 1652 of *Lecture Notes in Mathematics*.  
756 Springer Berlin, Heidelberg, 1997.
- 757 [Tod99] Stevo Todorcevic. Compact subsets of the first Baire class. *Journal of the American  
758 Mathematical Society*, 12(4):1179–1212, 1999.
- 759 [VC15] V. N. Vapnik and A. Ya. Chervonenkis. On the uniform convergence of relative frequen-  
760 cies of events to their probabilities. pages 11–30, 2015. Reprint of Theor. Probability  
761 Appl. **16** (1971), 264–280.